



CC patient suffering from a lysosomal storage disease. The present sequence  
 CC is used in the exemplification of the invention.

Sequence 928 AA;

Best Local Similarity

928;

Conservative

0;

Mismatches

0;

Indels

0;

Gaps

0;

Length

928;

Pred. No.

0;

Score

4907;

DB

4;

Match ID

ADD27812;

ID

ADD27812

Standard

protein

928 AA.

RESULT

2

XX

ID

ADD27812

standard

protein

928 AA;

XX

AC

ADD27812;

XX

DE

ADD27812;

XX

KW

ADD27812;

XX

OS

ADD27812;

XX

PR

ADD27812;

XX

PN

ADD27812;

XX

PD

ADD27812;

XX

PP

ADD27812;

XX

PR

ADD27812;

XX

PA

ADD27812;

XX

Db	301 YLMLDSAISQSOKODEDISASRFEDNEBLSRYSLSRSHAPWNRNIFTVNGQIPSWNL 360	DR	N-PSDB; AAD62491.
QY	361 DNPRTVTHQDFRNLSHPLPTSSPAIESHHRIGLSOKTYLNDVMGKDWPDDP 420	XX	
Db	361 DNPRTVTHQDFRNLSHPLPTSSPAIESHHRIGLSOKTYLNDVMGKDWPDDP 420	PT	Producing a high mannose glycoprotein for treating lysosomal storage disease, comprises culturing the Lectin resistant mammalian cell in the presence of deoxymannojirimycin and kifunensine.
QY	421 YSHSKGQKYLTPVPGCAEGPGSMWKDGYCDKACNSACWDGGCSCGSNSGSRVAG 480	XX	
Db	421 YSHSKGQKYLTPVPGCAEGPGSMWKDGYCDKACNSACWDGGCSCGSNSGSRVAG 480	PT	presence of deoxymannojirimycin and kifunensine.
QY	481 GGGTGSIGGHPMQFGGINSYCNOCANSWLADEFCDCQACNVISCGFDAGDCQDHF 540	XX	
Db	481 GGGTGSIGGHPMQFGGINSYCNOCANSWLADEFCDCQACNVISCGFDAGDCQDHF 540	PS	Claim 10; Page 15-18; 46pp; English.
QY	481 HELVKVILPNQTHYLTKGECLPYFSFAEVAKRGVGVAYSNPITRHASTANKWTHI 600	CC	The invention relates to a method for producing a high mannose glycoprotein. The method comprises: introducing and expressing a polynucleotide encoding a glycoprotein into a mammalian cell; culturing the cell in the presence of a lectin to obtain a lectin resistant cell; isolating the cell; culturing the cell in the presence of deoxymannojirimycin and kifunensine to inhibit glycosylation of the glycoprotein; and collecting the glycoprotein. The invention is useful in gene therapy. The method is useful for producing a high mannose glycoprotein in a complex carbohydrate deficient cell for treating lysosomal storage disease. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-phosphotransferase alpha subunit precursor protein.
Db	541 HELVKVILPNQTHYLTKGECLPYFSFAEVAKRGVGVAYSNPITRHASTANKWTHI 600	CC	
QY	541 HELVKVILPNQTHYLTKGECLPYFSFAEVAKRGVGVAYSNPITRHASTANKWTHI 600	CC	
Db	601 IMSGMMAITHNLFTONTNDBEFPQIIVTVDTRGPKUNSTACKGYENVSPTILP 660	CC	
QY	601 IMSGMMAITHNLFTONTNDBEFPQIIVTVDTRGPKUNSTACKGYENVSPTILP 660	CC	
Db	601 IMSGMMAITHNLFTONTNDBEFPQIIVTVDTRGPKUNSTACKGYENVSPTILP 660	XX	
QY	661 EAEFLFDTPKERPKKRHVNSTRAQEVKILPVNLSLPKAQLSNTDQLEH 720	XX	
Db	661 EAEFLFDTPKERPKKRHVNSTRAQEVKILPVNLSLPKAQLSNTDQLEH 720	XX	
QY	721 GDITLKGNLSSALLSFLMSQHAKTNQAITDINDSVAPEKQHKSILPNLG 780	QY	Sequence 928 AA;
Db	721 GDITLKGNLSSALLSFLMSQHAKTNQAITDINDSVAPEKQHKSILPNLG 780	QY	Query Match 100%; Score 4907; DB 7; Length 928; Best Local Similarity 100%; Pred. No. 0; Matches 928; Conservative 0; Mismatches 0; Indels 0; Caps 0; Polynucleotide 1 MLFKLILQRQTYTCIHSRYGLYVCFIGVYVTTAFAQFGEVWLEWSRQHVLPSYDN1 60
QY	781 VSRQLQITPPAVSVKRGHDQCNPLDLETTARRVETHTQKTTGGNVTKEKPSLIV 840	Db	1 MLFKLILQRQTYTCIHSRYGLYVCFIGVYVTTAFAQFGEVWLEWSRQHVLPSYDN1 60
Db	781 VSRQLQITPPAVSVKRGHDQCNPLDLETTARRVETHTQKTTGGNVTKEKPSLIV 840	QY	61 AGKSFRNRLCLPMPIDVYVTTWNGDLELKEQVQREMEFQKAREILGKNTTPK 120
QY	841 PLESQMTKEKKLTGKEKENSRSVRAENHIGTVTEVLGRKLUQHYSYPLPWKKYF 900	Db	61 AGKSFRNRLCLPMPIDVYVTTWNGDLELKEQVQREMEFQKAREILGKNTTPK 120
Db	841 PLESQMTKEKKLTGKEKENSRSVRAENHIGTVTEVLGRKLUQHYSYPLPWKKYF 900	QY	121 KSEKQECUCLTHCTKVPMLVLDPALPANTIKLKVPSLXPSFHSAASDIFPVAKPNPTNV 180
QY	901 QDLIDDEBSKLTKOLAYFIDSQKNGRQK 928	Db	121 KSEKQECUCLTHCTKVPMLVLDPALPANTIKLKVPSLXPSFHSAASDIFPVAKPNPTNV 180
Db	901 QDLIDDEBSKLTKOLAYFIDSQKNGRQK 928	QY	181 SVVFDSTKVEDAHSGLJKGNRSQTVWVRYLTDKEVPLVGLYMQDIAFLSCPPTKET 240
RESULT 3		Db	181 SVVFDSTKVEDAHSGLJKGNRSQTVWVRYLTDKEVPLVGLYMQDIAFLSCPPTKET 240
ID ABW01488		QY	241 NOLKTKLPENLSSKVYKLUOLYSEASVALKUNPKDFOELNKOTKNTIDCKELTSVA 300
ID ABW01488 standard; protein; 928 AA.		Db	241 NOLKTKLPENLSSKVYKLUOLYSEASVALKUNPKDFOELNKOTKNTIDCKELTSVA 300
AC ABW01488;		QY	301 YLMLDSAISQSOKODEDISASRFEDNEBLSRYSLSRSHAPWNRNIFTVNGQIPSWNL 360
DT 15-JAN-2004 (first entry)		Db	301 YLMLDSAISQSOKODEDISASRFEDNEBLSRYSLSRSHAPWNRNIFTVNGQIPSWNL 360
XX Human GlcNAc-phosphotransferase alpha subunit precursor protein.		QY	361 DNPRTVTHQDFRNLSHPLPTSSPAIESHHRIGLSOKTYLNDVMGKDWPDDP 420
XX Mannose glycoprotein; gene therapy; carbohydrate deficient cell; lysosomal storage disease; N-acetylglucosamine-1-phosphotransferase; XX		Db	361 DNPRTVTHQDFRNLSHPLPTSSPAIESHHRIGLSOKTYLNDVMGKDWPDDP 420
XX gastrointesinal; human; enzyme; lectin resistant cell; XX		QY	421 YSHSKGQKYLTPVPGCAEGPGSMWKDGYCDKACNSACWDGGCSCGSNSGSRVAG 480
XX deoxymannojirimycin; kifunensine; glycosylation inhibition. XX		Db	421 YSHSKGQKYLTPVPGCAEGPGSMWKDGYCDKACNSACWDGGCSCGSNSGSRVAG 480
OS Homo sapiens.		QY	481 GGGTGSIGGHPMQFGGINSYCNOCANSWLADEFCDCQACNVISCGFDAGDCQDHF 540
XX US2003124652-A1.		Db	481 GGGTGSIGGHPMQFGGINSYCNOCANSWLADEFCDCQACNVISCGFDAGDCQDHF 540
XX 03-JUL-2003.		QY	541 HELVKVILPNQTHYLTKGECLPYFSFAEVAKRGVGVAYSNPITRHASTANKWTHI 600
PD (NOVA-) NOVAYZME PHARM INC.		Db	541 HELVKVILPNQTHYLTKGECLPYFSFAEVAKRGVGVAYSNPITRHASTANKWTHI 600
XX Canfield WM;		QY	601 IMSGMMAITHNLFTONTNDBEFPQIIVTVDTRGPKUNSTACKGYENVSPTILP 660
DR WPI; 2003-810984/76.		Db	601 IMSGMMAITHNLFTONTNDBEFPQIIVTVDTRGPKUNSTACKGYENVSPTILP 660
XX		QY	661 EAEFLFDTPKERPKKRHVNSTRAQEVKILPVNLSLPKAQLSNTDQLEH 720

RESULT 4

ABR01537  
ID ABR01537 standard; protein; 928 AA.  
XX  
AC ABR01537;  
XX  
DT 15-JAN-2004 (first entry)  
DE Human GICNAc-phosphotransferase alpha subunit precursor protein.  
KW Glycoprotein; lectin; lysosomal storage disease; gastrointestinal; N-acetylglucosamine-1 phosphotransferase; enzyme; gene therapy; human.  
XX Homo sapiens.  
XX US2003124653-A1.  
XX  
PD 03-JUL-2003.  
PF 21-DEC-2001; 2001US-00023890.  
XX  
PR 21-DEC-2001; 2001US-00023890.  
XX  
PA (NOVA-) NOVAZYME PHARM INC.  
XX  
PI Canfield WM;  
XX  
WPI; 2003-8-0985/76.  
DR N-PSDB; AAD62650.

Producing a glycoprotein with reduced complex carbohydrates by culturing the lectin resistant mammalian cell expressing the glycoprotein for treating lysosomal storage disease.

PS Claim 10; Page 16-18; 46pp; English.

XX  
CC The present invention provides a method of producing a glycoprotein having reduced complex carbohydrates by culturing the lectin resistant mammalian cell expressing the glycoprotein. The method is useful for producing a glycoprotein with reduced complex carbohydrates for treating lysosomal storage disease. The present invention is also useful in gene therapy. The present sequence is human N-acetylglucosamine-1 (GICNAc)-phosphotransferase alpha subunit precursor protein

XX Sequence 928 AA;

Query Match 100.0%; Score 4907; DB 7; length 928;  
Best local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0; Matches 928; Conservative 0;

QY 1 MLFKLQLQQTYTCLSHRYGLYVCPFLGVVTTISAFOFGVVWLEWSRQHVLFDSDRNDI 60  
Db 1 MLFKLQLQQTYTCLSHRYGLYVCPFLGVVTTISAFOFGVVWLEWSRQHVLFDSDRNDI 60

Db 651 EAEIFEDIPKEKPKERHDNSTRAQEVKIPVNISLPPKQLSINTLQLEH 720  
ID AAE25290 standard; protein; 1256 AA.  
XX  
AC AAE25290;  
XX  
DT 30-OCT-2002 (first entry)  
DE Human nucleic acid-associated protein (NAAP-9).  
KW Human; nucleic acid-associated protein; NAAP-9; neurological disorder; arteriosclerosis; cancer; cell proliferative disorder; atherosclerosis;

QY 61 AGCSFRQNLCLPMPDVYUVTWNGTDIPLKLSQLQREQMBEQQKARREIGKNTTPTK 120  
Db 61 AGKSFRQNLCLPMPDVYUVTWNGTDIPLKLSQLQREQMBEQQKARREIGKNTTPTK 120  
ID 721 GDTIKGYNLSKSALLSPIMQSQHAKIKNOAIIITDETDNSLVAPQEKVHSILPNSLG 780  
Db 721 GDTIKGYNLSKSALLSPIMQSQHAKIKNOAIIITDETDNSLVAPQEKVHSILPNSLG 780  
QY 841 PLESQMTKEKKTKOKEKNSRMENAEHNIGHTEVILGRKLOHYTSYLGFLPWEKKYF 900  
Db 841 PLESQMTKEKKTKOKEKNSRMENAEHNIGHTEVILGRKLOHYTSYLGFLPWEKKYF 900  
QY 901 QDLDDEESLKTOLAYFTDSKNTGRQLK 928  
Db 901 QDLDDEESLKTOLAYFTDSKNTGRQLK 928

Db 121 KSEKQLECLLTHCKVPMVLDPALPANIITKDVPSLPSFSASIDFNVAKPNSTNV 180  
Db 121 KSEKQLECLLTHCKVPMVLDPALPANIITKDVPSLPSFSASIDFNVAKPNSTNV 180  
QY 181 SVVVFESTKDVEAHSGULKNSRQTWVRLGILTIDKEVPGVLMQDQLFLSCPPPFKET 240  
Db 181 SVVVFESTKDVEAHSGULKNSRQTWVRLGILTIDKEVPGVLMQDQLFLSCPPPFKET 240  
Db 241 NOLKTKLPNLSKVKLUQLYBASVALLKLNPKDQBLNQTKRUMTICKELTISPA 300  
Db 241 NOLKTKLPNLSKVKLUQLYBASVALLKLNPKDQBLNQTKRUMTICKELTISPA 300  
QY 301 YLMLDLSAJSOSKQDEDIASRFEEDNEFLYSLRSTERHAPAVRNFIYTNGQIPRWNL 360  
Db 301 YLMLDLSAJSOSKQDEDIASRFEEDNEFLYSLRSTERHAPAVRNFIYTNGQIPRWNL 360  
QY 361 DNPRVITVHODPYRNTSHLPTSSPAESEHTRIEGSIQYIYNDVMGKDVPDPDF 420  
Db 361 DNPRVITVHODPYRNTSHLPTSSPAESEHTRIEGSIQYIYNDVMGKDVPDPDF 420  
QY 421 YSHKGQKYLTPWPNAGCGSWIKDGCDKACNSACWDGGDCSGNSGSRSRYTAG 480  
Db 421 YSHKGQKYLTPWPNAGCGSWIKDGCDKACNSACWDGGDCSGNSGSRSRYTAG 480  
QY 481 GGGTGSGIGVGHMOPGGGINSYICNOGANSWADXFCDQACQANVUSCGFDAGCQDHF 540  
Db 481 GGGTGSGIGVGHMOPGGGINSYICNOGANSWADXFCDQACQANVUSCGFDAGCQDHF 540  
QY 541 HELYKVILLPNQTHYIIKMGCECPYPSFAEVAKRGVEGAYSNDNPTRHASTANKWTHL 600  
Db 541 HELYKVILLPNQTHYIIKMGCECPYPSFAEVAKRGVEGAYSNDNPTRHASTANKWTHL 600  
QY 601 IMHEGMNATTIHTNLTIFONTNDDEBFKQIQTVEUDTRGPKUNSTAOKGYENIVSPITLLP 660  
Db 601 IMHEGMNATTIHTNLTIFONTNDDEBFKQIQTVEUDTRGPKUNSTAOKGYENIVSPITLLP 660  
QY 661 EAEIFEDIPKEKPKERHDNSTRAQEVKIPVNISLPPKQLSINTLQLEH 720  
Db 661 EAEIFEDIPKEKPKERHDNSTRAQEVKIPVNISLPPKQLSINTLQLEH 720  
QY 721 GDTIKGYNLSKSALLSPIMQSQHAKIKNOAIIITDETDNSLVAPQEKVHSILPNSLG 780  
Db 721 GDTIKGYNLSKSALLSPIMQSQHAKIKNOAIIITDETDNSLVAPQEKVHSILPNSLG 780  
QY 781 VSERLQRLTPAVSKVNGHDOQNPPDLETTARFRVETHIQKIGNTGKPPSLIV 840  
Db 781 VSERLQRLTPAVSKVNGHDOQNPPDLETTARFRVETHIQKIGNTGKPPSLIV 840  
QY 841 PLESQMTKEKKTKOKEKNSRMENAEHNIGHTEVILGRKLOHYTSYLGFLPWEKKYF 900  
Db 841 PLESQMTKEKKTKOKEKNSRMENAEHNIGHTEVILGRKLOHYTSYLGFLPWEKKYF 900  
QY 901 QDLDDEESLKTOLAYFTDSKNTGRQLK 928  
Db 901 QDLDDEESLKTOLAYFTDSKNTGRQLK 928

RESULT 5

AAE25290  
ID AAE25290 standard; protein; 1256 AA.  
XX  
AC AAE25290;  
XX  
DT 30-OCT-2002 (first entry)  
DE Human nucleic acid-associated protein (NAAP-9).  
KW Human; nucleic acid-associated protein; NAAP-9; neurological disorder; arteriosclerosis; cancer; cell proliferative disorder; atherosclerosis;

KW	lymphoma; epilepsy; Alzheimer's disease; developmental; anticonvulsant;	Db	121 KSEKOLECLLTHCIKVPMLVLDPALPANITLKOLPLSYPSFHSAIDIFNVAKPKNPSTNV 180
KW	autoimmune disorder; AIDS; allergy; anaemia; stroke; malaria; leishmania;	Qy	181 SVVVFDSKTDVKEDAHSGILKGNSRQTVARGYJTTDKEVPGLYMIMQDIAFLSCFPPTKEF 240
KW	gene therapy; nootropic; neuroprotective; cerebroprotective; virucide;	Db	181 SVVVFDSKTDVKEDAHSGILKGNSRQTVARGYJTTDKEVPGLYMIMQDIAFLSCFPPTKEF 240
KW	immuno suppressive; protozoacide; antimicrobial.	Qy	241 NOLKTKLPENLSSKVYKLYSRSVALKKLNPKDQEELNKOTKNTICKELTSP 300
OS	Homo sapiens.	Db	241 NOLKTKLPENLSSKVYKLYSRSVALKKLNPKDQEELNKOTKNTICKELTSP 300
XX		Qy	301 YLLWDLSATISQSQKDEDISASRFEDNEILYRSIRHAPMTNIFTNGQIPSMNL 360
FT	Key	Location/Qualifiers	301 YLLWDLSATISQSQKDEDISASRFEDNEILYRSIRHAPMTNIFTNGQIPSMNL 360
FT	Domain	448 . 469 /note= "Notch domain"	301 YLLWDLSATISQSQKDEDISASRFEDNEILYRSIRHAPMTNIFTNGQIPSMNL 360
FT	Domain	500 . 536 /note= "Notch domain"	301 YLLWDLSATISQSQKDEDISASRFEDNEILYRSIRHAPMTNIFTNGQIPSMNL 360
FT	Domain	1018 . 1030 /note= "EF-hand calcium-binding domain"	301 YLLWDLSATISQSQKDEDISASRFEDNEILYRSIRHAPMTNIFTNGQIPSMNL 360
XX	WO200250279-A2.	Qy	361 DNPRVITVHQDFRNLSLHPLTPSSPAATESHRIEJSQKTYIINDVMRDKDVWDDF 420
PN		Db	361 DNPRVITVHQDFRNLSLHPLTPSSPAATESHRIEJSQKTYIINDVMRDKDVWDDF 420
XX	PD	PD	361 DNPRVITVHQDFRNLSLHPLTPSSPAATESHRIEJSQKTYIINDVMRDKDVWDDF 420
XX	19-DEC-2-2001; 2001WO-US050256.	Qy	421 YSHRKQKYLTMVPUVNAEGPGPSWIKGYCDKACCHNSACWDGGCGSGNSGSRVIA 480
XX		Db	421 YSHRKQKYLTMVPUVNAEGPGPSWIKGYCDKACCHNSACWDGGCGSGNSGSRVIA 480
PR	21-DEC-2-2000; 2000US-0257714P.	Qy	481 GGGTGSISGVQWPWFGGGINSVSYCNOCANSILADKFCQDNVSCGFBGDCGDHF 540
PR	05-JAN-2-2001; 2001US-0262102P.	Db	481 GGGTGSISGVQWPWFGGGINSVSYCNOCANSILADKFCQDNVSCGFBGDCGDHF 540
PR	16-JAN-2-2001; 2001US-0262102P.	Qy	541 HELKUWILPQNQTHYTIKGECIPLYFSPAEVARQGVGAYASNPIIHASTANKWKITH 600
PR	23-JAN-2-2001; 2001US-0253823P.	Db	541 HELKUWILPQNQTHYTIKGECIPLYFSPAEVARQGVGAYASNPIIHASTANKWKITH 600
PR	02-FEB-2-2001; 2001US-0266088P.	Qy	541 HELKUWILPQNQTHYTIKGECIPLYFSPAEVARQGVGAYASNPIIHASTANKWKITH 600
PR	29-OCT-2-2001; 2001US-0348442P.	Db	541 HELKUWILPQNQTHYTIKGECIPLYFSPAEVARQGVGAYASNPIIHASTANKWKITH 600
XX		Qy	541 HELKUWILPQNQTHYTIKGECIPLYFSPAEVARQGVGAYASNPIIHASTANKWKITH 600
PA	(INCY.) INCITE GENOMICS INC.	Db	541 HELKUWILPQNQTHYTIKGECIPLYFSPAEVARQGVGAYASNPIIHASTANKWKITH 600
XX		Qy	541 HELKUWILPQNQTHYTIKGECIPLYFSPAEVARQGVGAYASNPIIHASTANKWKITH 600
PI	Baughn MR, Lu Y, Arvizu C, Ramkumar J, Yao MG, Policky JL;	Qy	601 IMESGMNATIHEMLTQNTNDBEFKMOTIVETVDTRGPKLNSTAQKGYENLSPITL 660
PI	Walia NK, Tribouley RM, Yue H, Batra S, Ding L, Ial PG;	Db	601 IMESGMNATIHEMLTQNTNDBEFKMOTIVETVDTRGPKLNSTAQKGYENLSPITL 660
PI	Borowsky ML, Lu DAM, Gandhi AR, Griffin JA, Xu Y, Azimzai Y;	Qy	661 EAEILPEDIKPKEKPPKGRADNSTRAQEEKPKJUNISLPLQDQLSNTDQLEH 720
PI	Gietzen KJ, Tang YT, Warren BA, Mason PM, Burkord N, Bafaila AUN;	Db	661 EAEILPEDIKPKEKPPKGRADNSTRAQEEKPKJUNISLPLQDQLSNTDQLEH 720
PI	Lee EA, Yang J, Govard AE, Emerling BM, Marquis JP, Lee SY;	Qy	721 GDITLKGYNLSKALLRSLIMOSOHAALKQNQAITDENDSLVAPQRQVHSILPNSLG 780
PI	Swarnakar A, Reddy R;	Db	721 GDITLKGYNLSKALLRSLIMOSOHAALKQNQAITDENDSLVAPQRQVHSILPNSLG 780
DR	WPI: 2002-519887-55.	Qy	781 VSERLQLRIFPAPSVKVGHDQONPPLDETTARFVETHTQKTCGNVTKEPKPLIV 840
DR	N-PSDB; AAD41199.	Db	781 VSERLQLRIFPAPSVKVGHDQONPPLDETTARFVETHTQKTCGNVTKEPKPLIV 840
XX		Qy	841 PLESQMTTEKKITGKEKNSRMEAAEHIGVTEVJGKQHYTISYLGEPWEEKYF 900
PT	Nucleic acid associated proteins and nucleic acids for diagnosing,	Db	841 PLESQMTTEKKITGKEKNSRMEAAEHIGVTEVJGKQHYTISYLGEPWEEKYF 900
PT	treating and preventing cell proliferative (e.g. cancers), neurological	Qy	901 QDLDDEBLSIKLQAYFDKSNQGRLK 928
PT	(e.g. epilepsy or stroke) or autoimmune disorders (e.g. AIDS).	Db	901 QDLDDEBLSIKLQAYFDKSNQGRLK 928
XX	Claim 64; Page 163-165, 193pp; English.	Qy	901 QDLDDEBLSIKLQAYFDKSNQGRLK 928
CC	The invention relates to nucleic acid-associated proteins (NAAP) and	Db	901 QDLDDEBLSIKLQAYFDKSNQGRLK 928
CC	nucleic acids. The nucleic acid and amino acid sequences are useful for	Qy	901 QDLDDEBLSIKLQAYFDKSNQGRLK 928
CC	diagnosing, treating and preventing cell proliferative e.g.	Db	901 QDLDDEBLSIKLQAYFDKSNQGRLK 928
CC	arteriosclerosis, atherosclerosis, lymphoma or cancers), neurological	Qy	901 QDLDDEBLSIKLQAYFDKSNQGRLK 928
CC	(e.g. epilepsy, Alzheimer's disease or stroke), developmental, and	Db	901 QDLDDEBLSIKLQAYFDKSNQGRLK 928
CC	autoimmune disorders (e.g. AIDS, allergies, or anaemia) or infections	Qy	901 QDLDDEBLSIKLQAYFDKSNQGRLK 928
CC	(e.g. malaria, or leishmania), as well as in assessing the effects of	Db	901 QDLDDEBLSIKLQAYFDKSNQGRLK 928
CC	exogenous compound on the expression of nucleic acid and amino acid	Qy	901 QDLDDEBLSIKLQAYFDKSNQGRLK 928
CC	sequences of nucleic acid associated proteins. The invention is useful in	Db	901 QDLDDEBLSIKLQAYFDKSNQGRLK 928
CC	gene therapy. The present sequence is human NAAP-9	Qy	901 QDLDDEBLSIKLQAYFDKSNQGRLK 928
XX	Sequence 1256 AA;	SO	RESULT 6
SO	Query Match: Score 4896; DB 5; Length 1256;	ID	ADD27810
SO	Best Local Similarity: 99.8%; Score: 4896; DB: 5; Length: 1256;	XX	ADD27810 standard; protein: 1199 AA.
SO	Matches: 926; Conservative: 1; Mismatches: 0; Indels: 0; Gaps: 0;	AC	ADD27810;
Oy	1 MIFPKLQLRQTYCILSHRYGLYVCFIGVWVITVAFAQFGVEVLEWRDQYHFLFSYRDNT 60	XX	15-JAN-2004 (first entry)
Db	1 MIFPKLQLRQTYCILSHRYGLYVCFIGVWVITVAFAQFGVEVLEWRDQYHFLFSYRDNT 60	XX	Soluble human GlcNAc-phosphotransferase.
Db	61 AGKSFQRLCQMPIDVWVTTWNGTDLKELQLOREQMEOKAMRELGKNTPEK 120	XX	human; protein phosphorylation; soluble GlcNAc-phosphotransferase;
Db	61 AGKSFQRLCQMPIDVWVTTWNGTDLKELQLOREQMEOKAMRELGKNTPEK 120	XX	UDP-N-acetylglucosamine; lysosomal storage disease; Fabry's disease; enzyme.
Oy	121 KSEKOLECLLTHCIKVPMLVLDPALPANITLKOLPLSYPSFHSAIDIFNVAKPKNPSTNV 180	OS	Synthetic.
PN	US2003119088-A1.	XX	Homo sapiens.

PD	26-JUN-2003.	QY	645	AQKGSYENLVSPIITLPEAIIIFDIPKKEPKPFKRDVNSTRRAOBVKTPLNVLSP
XX			635	
XX	21-DEC-2001; 2001US-00023888.	Db	635	AQKGSYENLVSPIITLPEAIIIFDIPKKEPKPFKRDVNSTRRAOBVKTPLNVLSP
PR	21-DEC-2001; 2001US-00023888.	QY	705	KDAQLSLATIDLOLEHGDTIKSYNLSKALRSFLANSQHAKIKNGAIIITDETDNSVA
XX	(NOVA-) NOVAZYME PHARM INC.	Db	705	764
PA	Canfield W, Kudo M;	QY	695	KDAQLSLATIDLOLEHGDTIKSYNLSKALRSFLANSQHAKIKNGAIIITDETDNSVA
XX	WPI; 2003-801323/75.	Db	695	754
DR	N-PSDB; ADB27809.	QY	765	POEKVHKSILPUSLPGVBERLQIPTAVSKVNGHQGONPLDUTTARVREHTOK
XX	PT	Db	765	824
PT	lysosomal storage disease e.g. Fabry's disease by contacting the protein with a soluble GlcNAc-phosphotransferase and producing a phosphorylated protein.	QY	755	POEQVHKSTILPUSLPGVBERLQIPTAVSKVNGHQGONPLDUTTARVREHTOK
CC	e.g. Fabry's disease. The present sequence represents the amino acid sequence of soluble human GlcNAc-phosphotransferase.	Db	755	814
CC	Sequence 1199 AA;	QY	825	TIGANUTKEPPSILVPLSQMMKEKITGKEKNSMEENAHICTEVILGRKLYH
XX	SQ	Db	825	884
PS	Claim 3; SEQ ID NO 2; 55pp; English.	QY	815	TIGANUTKEPPSILVPLSQMMKEKITGKEKNSMEENAHICTEVILGRKLYH
XX	The invention relates to a method of phosphorylating a protein comprising contacting the protein with a soluble GlcNAc-phosphotransferase (UDP-N-acetylglucosamine) and producing a phosphorylated protein. The method is useful for treating a patient suffering from a lysosomal storage disease e.g. Fabry's disease. The present sequence represents the amino acid sequence of soluble human GlcNAc-phosphotransferase.	Db	815	874
CC	Sequence 1199 AA;	QY	885	TDVSLGLEWERKKIYQDLDEBESLKQLAYFTDSKRAR
XX	SQ	Db	885	915
Query Match	94.4%; Score 4630; DB 7; Length 1199;	RESULT	7	
Best Local Similarity	99.3%; Pred. No. 0;	ID	ABW01487	
Matches	875; Conservative 2; Mismatches 4; Indels 0; Gaps 0;	ID	ABW01487 standard; protein; 1199 AA.	
Qy	45 SDDOYHFLFDSTRDNAGSKSFRNLCLPMPDIDVYTWNGSDLELLKELQVREONEEEQ	XX	ABW01487;	
Db	35 SRDQHVLPFDSTRDNAGSKSFRNLCLPMPDIDVYTWNGSDLELLKELQVREONEEEQ	XX	AC	
Qy	95 KAMREILGKNTTRPTKSEKOCELCIITHCIKVPMLDQDPAINTLKDQPSLSPSFSA	XX	XX	
Db	95 KAMREILGKNTTRPTKSEKOCELCIITHCIKVPMLDQDPAINTLKDQPSLSPSFSA	XX	XX	
Qy	165 SDIFNVAKPKRNSITNWSVNUFESTKVEDAHSGLKNSRQTRWRCYLTDKEVGLVM	XX	DE	
Db	155 SDIFNVAKPKRNSITNWSVNUFESTKVEDAHSGLKNSRQTRWRCYLTDKEVGLVM	XX	DE	
Qy	225 QDLAFLSGFPPTKETNQIQLKTPKLNNSKVKUOJYSEAVALIQLNNPKDFQELANKQT	XX	KW	
Db	225 QDLAFLSGFPPTKETNQIQLKTPKLNNSKVKUOJYSEAVALIQLNNPKDFQELANKQT	XX	KW	
Qy	215 QDLAFLSGFPPTKETNQIQLKTPKLNNSKVKUOJYSEAVALIQLNNPKDFQELANKQT	XX	KW	
Db	215 QDLAFLSGFPPTKETNQIQLKTPKLNNSKVKUOJYSEAVALIQLNNPKDFQELANKQT	XX	KW	
Qy	285 KKMTNDCKELTISPAVILWIDSAISOKBEDISSRFEDNBELRSLISTERIAWPWR	XX	KW	
Db	275 KKMTNDCKELTISPAVILWIDSAISOKBEDISSRFEDNBELRSLISTERIAWPWR	XX	KW	
Qy	345 NIFIVTNGQI PSMKLNLQNPVITVHDVFNISHPTFSSPAIISHTIRREGISOKFV	XX	KW	
Db	335 NIFIVTNGQI PSMKLNLQNPVITVHDVFNISHPTFSSPAIISHTIRREGISOKFV	XX	KW	
Qy	405 LNDWWFCKDWPPDDFISHSKGOKVYLTMPVNCAGCGPGSWIKOQYCDKACNNACDWD	XX	PA	
Db	405 LNDWWFCKDWPPDDFISHSKGOKVYLTMPVNCAGCGPGSWIKOQYCDKACNNACDWD	XX	PA	
Qy	395 LNDWWFCKDWPPDDFISHSKGOKVYLTMPVNCAGCGPGSWIKOQYCDKACNNACDWD	XX	PA	
Db	395 LNDWWFCKDWPPDDFISHSKGOKVYLTMPVNCAGCGPGSWIKOQYCDKACNNACDWD	XX	PA	
Qy	464 GGDGCGGGSGGSRYTAGGAGGTGSGTIGWGPWQEGGINMSVSYCNQGANSWIAKFCQACN	XX	CC	
Db	455 GGDGCGGGSGGSRYTAGGAGGTGSGTIGWGPWQEGGINMSVSYCNQGANSWIAKFCQACN	XX	CC	
Qy	525 VLSCGFDGDCQDHFLBLKYLKVILLPNQTHYIIPKSCCLPPTFAWAKRVEGAYSDNP	XX	CC	
Db	515 VLSCGFDGDCQDHFLBLKYLKVILLPNQTHYIIPKSCCLPPTFAWAKRVEGAYSDNP	XX	CC	
Qy	585 TIRHASTANKWKHLHMGKNAITFHNLQFONTDEEKFQKQIVETVDTREGKLNST	XX	CC	
Db	575 TIRHASTANKWKHLHMGKNAITFHNLQFONTDEEKFQKQIVETVDTREGKLNST	XX	CC	
Qy	644	Sequence 1199 AA;	QY	
Db	634	Sequence 1199 AA;	QY	

Query Match 94.4%; Score 4630; DB 7; Length 1199;  
 Best Local Similarity 99.3%; Pred. No. 0; Mismatches 2; Indels 0; Gaps 0;  
 Matches 875; Conservative 4; OS Unidentified.

QY 45 SRDQYHVLFDSDRDNIAKSFONRLCLPMPIDVWVWVNGIDJELIKELOVREGEQQEQQ 104  
 DB 35 SDQYHVLFDSRDNIAKSFONRLCLPMPIDVWVWVNGIDJELIKELOVREGEQQEQQ 94  
 QY 105 KAMBLIGKNTTEPTKSKEKQLECLLTHICKRVPMLVDPALPANTIKDPSLYSPHSA 164  
 DB 95 KAMBLIGKNTTEPTKSKEKQLECLLTHICKRVPMLVDPALPANTIKDPSLYSPHSA 154  
 QY 165 SDIFNVAKPKNPSTNSVWVFUDSTKVEDAHSGILKNSRQTWVGRGLTIDKEVGVLM 224  
 DB 155 SDIFNVAKPKNPSTNSVWVFUDSTKVEDAHSGILKNSRQTWVGRGLTIDKEVGVLM 214  
 QY 225 QDIAFLSGFPPPTKETNOIJKTKEPNTSSKVKLQIOLSEASVALKANNPKOFOELNKQT 284  
 DB 215 QDIAFLSGFPPPTKETNOIJKTKEPNTSSKVKLQIOLSEASVALKANNPKOFOELNKQT 274  
 QY 285 KAMBLIGKNTTEPTKSKEKQLECLLTHICKRVPMLVDPALPANTIKDPSLYSPHSA 344  
 DB 275 KAMBLIGKNTTEPTKSKEKQLECLLTHICKRVPMLVDPALPANTIKDPSLYSPHSA 334  
 QY 345 NIFITVNGQIPSWNLNDPRTVTHOFRNLSHLPTSPAESHRIGLSCPY 404  
 DB 335 NIFITVNGQIPSWNLNDPRTVTHOFRNLSHLPTSPAESHRIGLSCPY 394  
 QY 405 LNDDMFGKDVWDDPFYSHSKGQKVLTWVPGCPSWIKOGYCDKACNNACDW 464  
 DB 395 LNDDMFGKDVWDDPFYSHSKGQKVLTWVPGCPSWIKOGYCDKACNNACDW 454  
 QY 465 GGDSGSNSGGSRYTAGGGTGSTGVGHPWQFGGINSVSYNCOGCANSLADKFCDAQN 524  
 DB 455 GGDSGSNSGGSRYTAGGGTGSTGVGHPWQFGGINSVSYNCOGCANSLADKFCDAQN 514  
 QY 525 VUSCGFDAQCGQDPFHLLKUILLPQHTYIPKGRCJPLPSAEVAKROVEGAYSDNP 584  
 DB 515 VUSCGFDAQCGQDPFHLLKUILLPQHTYIPKGRCJPLPSAEVAKROVEGAYSDNP 574  
 QY 585 IIRHASTANKWKTITHLIMSGNATTHFLNLFQNTDEERKQIVTEVDTRGPKLNST 644  
 DB 575 IIRHASTANKWKTITHLIMSGNATTHFLNLFQNTDEERKQIVTEVDTRGPKLNST 634  
 QY 645 AOKGYENIVSPTTLPPEALBFDIPEKRRPKPKRDVNTRRAEBCVKIPVLTSLP 704  
 DB 635 AOKGYENIVSPTTLPPEALBFDIPEKRRPKPKRDVNTRRAEBCVKIPVLTSLP 694  
 QY 705 KDKQSLNTLDIQLEHGTITLGKYNUSKALLRSFLMSQRHKQQAITDETDSLVA 764  
 DB 695 KDKQSLNTLDIQLEHGTITLGKYNUSKALLRSFLMSQRHKQQAITDETDSLVA 754  
 QY 765 P0EKQVHKSILPNLGLYSERLQLTPAVSVKNGHDQGPPLDETTARFRVHTQK 824  
 DB 755 P0EKQVHKSILPNLGLYSERLQLTPAVSVKNGHDQGPPLDETTARFRVHTQK 814  
 QY 825 TIGGNVTKKEPKPLIVPLSOMTKEKKTGKEKENSMEENAHIGVTEVLLGRKQHY 884  
 DB 815 TIGGNVTKKEPKPLIVPLSOMTKEKKTGKEKENSMEENAHIGVTEVLLGRKQHY 874  
 QY 885 TDSYLGFLFWEKKYFQDLIDDEBESLKTYQLAYFTDSKNTGR 925  
 DB 875 TDYLGFLFWEKKYFQDLIDDEBESLKTYQLAYFTDSKNTGR 915  
 RESULT 8  
 ABW01536 standard, protein: 1199 AA.  
 ID ABW01536  
 XX AC ABW01536;  
 DT 15-JAN-2004 (first entry)

XX DE N-acetylglucosamine-1(GlcNAc)-phosphotransferase protein.  
 XX KW Glycoprotein; lectin; lysosomal storage disease; gastrointestinal; N-acetylglucosamine-1-phosphotransferase; gene therapy; enzyme.  
 XX OS Unidentified.

XX PN US2003124653.A1.

XX PD 03-JUL-2003.

XX PF 21-DEC-2001; 2001US-00023890.

XX PR 21-DEC-2001; 2001US-00023890.

XX PA (NOVA-) NOVAZINE PHARM INC.

XX PI Camfield WM;

XX DR N-PSDB; RRD62649.

XX CC Producing a glycoprotein with reduced complex carbohydrates by culturing  
 PT the lectin resistant mammalian cell expressing the glycoprotein for  
 CC producing a glycoprotein with reduced complex carbohydrates for treating  
 PT lysosomal storage disease. The present invention is also useful in gene  
 CC therapy. The present sequence is N-acetylglucosamine (GlcNAc)-  
 CC phosphotransferase protein.

PS Claim 8; Page 10-13; 46pp; English.

The present invention provides a method of producing a glycoprotein having reduced complex carbohydrates by culturing the lectin resistant mammalian cell expressing the glycoprotein. The method is useful for producing a glycoprotein with reduced complex carbohydrates for treating lysosomal storage disease. The present invention is also useful in gene therapy. The present sequence is N-acetylglucosamine (GlcNAc)-phosphotransferase protein.

Sequence 1199 AA;

Query Match 94.4%; Score 4630; DB 7; Length 1199;  
 Best Local Similarity 99.3%; Pred. No. 0; Mismatches 2; Indels 0; Gaps 0;  
 Matches 875; Conservative 4; OS Unidentified.

QY 45 SRDQYHVLFDSDRDNIAKSFONRLCLPMPIDVWVWVNGIDJELIKELOVREGEQQEQQ 104  
 DB 35 SDQYHVLFDSRDNIAKSFONRLCLPMPIDVWVWVNGIDJELIKELOVREGEQQEQQ 94  
 QY 105 KAMBLIGKNTTEPTKSKEKQLECLLTHICKRVPMLVDPALPANTIKDPSLYSPHSA 164  
 DB 95 KAMBLIGKNTTEPTKSKEKQLECLLTHICKRVPMLVDPALPANTIKDPSLYSPHSA 154  
 QY 165 SDIFNVAKPKNPSTNSVWVFUDSTKVEDAHSGILKNSRQTWVGRGLTIDKEVGVLM 224  
 DB 155 SDIFNVAKPKNPSTNSVWVFUDSTKVEDAHSGILKNSRQTWVGRGLTIDKEVGVLM 214  
 QY 225 QDIAFLSGFPPPTKETNOIJKTKEPNTSSKVKLQIOLSEASVALKANNPKOFOELNKQT 284  
 DB 215 QDIAFLSGFPPPTKETNOIJKTKEPNTSSKVKLQIOLSEASVALKANNPKOFOELNKQT 274  
 QY 285 KAMBLIGKNTTEPTKSKEKQLECLLTHICKRVPMLVDPALPANTIKDPSLYSPHSA 344  
 DB 275 KAMBLIGKNTTEPTKSKEKQLECLLTHICKRVPMLVDPALPANTIKDPSLYSPHSA 334  
 QY 345 NIFITVNGQIPSWNLNDPRTVTHOFRNLSHLPTSPAESHRIGLSCPY 404  
 DB 335 NIFITVNGQIPSWNLNDPRTVTHOFRNLSHLPTSPAESHRIGLSCPY 394  
 QY 405 LNDDMFGKDVWDDPFYSHSKGQKVLTWVPGCPSWIKDGCDKACNNACDW 464  
 DB 395 LNDDMFGKDVWDDPFYSHSKGQKVLTWVPGCPSWIKDGCDKACNNACDW 454  
 QY 465 GGDSGSNSGGSRYTAGGGTGSTGVGHPWQFGGINSVSYNCOGCANSLADKFCDAQN 524

Db	455 GGDGSGNSGGSRSYIAGGGGTGSTGQVQPMQFGGINSVSYCNGCANSWLAQFCDOACN	PR	05-JAN-2001; 2001US-0260081P.	
QY	525 VLSCGFDAAGDCGODPHFLHLYKVILPONGTHYIIPKGECILPVPSAEVAKRGVEGAVSDNP	PR	16-JAN-2001; 2001US-0262302P.	
Db	515 VLSQGFDAAGDCGODPHFLHLYKVILPONGTHYIIPKGECILPVPSAEVAKRGVEGAVSDNP	PR	23-JAN-2001; 2001US-0263823P.	
QY	585 IIRIASIAKWKTHLHMGMTATHFNLFRONTDEBFMQITVEDREGPKNST	PR	02-FEB-2001; 2001US-0266088P.	
Db	635 AOKGYENLVSPIITPEALIFEDIPKRPFKRPFKRAUDNSURRAQEVKPLVN-SLIP	PR	29-OCT-2001; 2001US-0348442P.	
QY	575 IIRIASIAKWKTHLHMGMTATHFNLFRONTDEBFMQITVEDREGPKNST	PA	(INCYT-) INCYTE GENOMICS INC.	
Db	645 AOKGYENLVSPIITPEALIFEDIPKRPFKRPFKRAUDNSURRAQEVKPLVN-SLIP	PA	XX	
Db	695 KDAQSLINTLQDLOEHDGDLTCKYKPNLSSALRSFLMSQHAKKQAIIDETNSLVA	PA	XX	
QY	705 POEKQVHKISLPLNSLGVBRQLRUTFPASVVKNGHQGQNRPDLBTTARPRVHTQ	PA	XX	
Db	755 PQEQQVHKISLPLNSLGVBRQLRUTFPASVVKNGHQGQNRPDLBTTARPRVHTQ	PA	XX	
QY	825 TIGENNTKEKPSSLIVPESOMKKEKKTGKEKENSMEENAHNGTETLGRKQHY	PS	Walla NK, Tibouley KM, Yue H, Batra S, Ding L, Lal PG;	
Db	815 TIGENNTKEKPSSLIVPESOMKKEKKTGKEKENSMEENAHNGTETLGRKQHY	PS	Borowsky ML, Lu DAW, Gandhi AR, Griffin JA, Xu Y, Azimzai Y; Giezen KJ, Tang YT, Warren BA, Mason PM, Burford N, Hatalia AJA; Lee EA, Yang J, Goryad AE, Emerging BM, Marquis JP, Lee SY;	
QY	885 TDSTLGFLPWEKKEYFQDIDREBESIKQAYLAVFTDSKTR 925	PI	Swarnkar A, Reddy R,	
Db	875 TDSTLGFLPWEKKEYFQDIDREBESIKQAYLAVFTDSKTR 915	PI	DR	
RESULT 9	AAB25294	XX	WPI: 2002-519887/55.	
ID	AAB25294 standard; protein; 1196 AA.	XX	N-PDB; Aad41203.	
AC	AAB25294;	XX		
DT	30-OCT-2002 (first entry)	XX		
XX	Human nucleic acid-associated protein (NAAP-13).	XX		
XX	Human; nucleic acid-associated protein; NAAP-13; neurological disorder; arteriosclerosis; cancer; cell proliferative disorder; atherosclerosis; lymphoma; epilepsy; Alzheimer's disease; developmental; anticonvulsant; autoimmune disorder; AIDS; allergy; anaemia; stroke; malaria; leishmania; gene therapy; nootropic; neuroprotective; cerebroprotective; virucide; immunosuppressive; protozoicide; antimicrobial.	XX		
XX	Homo sapiens.	XX		
FT	Key	Location/Qualifiers		
FT	Peptide	1. .34		
FT	Domain	/Label= Signal_peptide		
FT	Domain	1..19		
FT	Domain	/note= "Cytosolic domain"		
FT	Domain	20..42		
FT	Protein	/note= "Transmembrane domain"		
FT	Protein	35..1196		
FT	Domain	/note= "Mature human NAAP-13"		
FT	Domain	43..1152		
FT	Domain	/note= "Non-cytosolic domain"		
FT	Domain	1153..1175		
FT	Domain	/note= "Transmembrane domain"		
FT	Domain	1176..1196		
FT	Domain	/note= "Cytosolic domain"		
FT	Domain	W0200250279-A2.		
PN	27-JUN-2002.	XX		
PF	19-DEC-2001; 2001WO-US050256.	XX		
PR	21-DEC-2000; 2000US-0257714P.	XX		
Query Match	93.4%; Score 4500; DB 5; Length 1196; Best Local Similarity	91.7%; Pred. No. 0; No. 0; Mismatches 0; Indels 60; Gaps 1; Matches 867; Conservative 1; MisMatches 0; InDelS 60; Gaps 1;	PR	05-JAN-2001; 2001US-0260081P.
QY	1 MLFKLQLOROTYTCISRYGLYVCFLGTVTTSAFPGEVVLEWSRDQYHFLDSRDNI	Db	16-JAN-2001; 2001US-0262302P.	
QY	1 MLFKLQLOROTYTCISRYGLYVCFLGTVTTSAFPGEVVLEWSRDQYHFLDSRDNI	Db	23-JAN-2001; 2001US-0263823P.	
QY	61 AGKSFONRLCLPMPIDVWVTTWNGTDLBLLKELQOREQMBEOKAMREILGNTTEPK	Db	02-FEB-2001; 2001US-0266088P.	
QY	61 AGKSFONRLCLPMPIDVWVTTWNGTDLBLLKELQOREQMBEOKAMREILGNTTEPK	Db	29-OCT-2001; 2001US-0348442P.	
QY	121 KSEKQECUCLTECKVPMVLDPALPANITKDVSPLYSRHSASIDPNVKPKNSTV	Db	(INCYT-) INCYTE GENOMICS INC.	
QY	121 KSEKQECUCLTECKVPMVLDPALPANITKDVSPLYSRHSASIDPNVKPKNSTV	Db	XX	
QY	181 SVVFFDSTKVEDAHSGLIKGNQRSTWGRGYLTDDKEVPGJLMQDPLFSLGPPPKET	Db	XX	
QY	181 SVVFFDSTKVEDAHSGLIKGNQRSTWGRGYLTDDKEVPGJLMQDPLFSLGPPPKET	Db	XX	
QY	241 NOLKTKLBNLSKVKLQLQSSEASVALLKLNPKOFOELNKQTKMTDGKELTISPA	Db	XX	
QY	241 NOLKTKLBNLSKVKLQLQSSEASVALLKLNPKOFOELNKQTKMTDGKELTISPA	Db	XX	
QY	301 YLFLWDSLATSQSKDEDISASRBDNFBLRVLRSRBRHAFWVRNTPITVNGQISWNL	Db	XX	
QY	301 YLFLWDSLATSQSKDEDISASRBDNFBLRVLRSRBRHAFWVRNTPITVNGQISWNL	Db	XX	
QY	361 DNPRVTIVTHQDFRNLSPFSSAIESHRIGISQSPYIYNDVMGKDWPDPDF	Db	XX	
QY	361 DNPRVTIVTHQDFRNLSPFSSAIESHRIGISQSPYIYNDVMGKDWPDPDF	Db	XX	
QY	421 YSHSGQKQVLTWPMVPCAGEGPGSMKDGICDKAQNNSACWDGDCSGNSGGSYTAG	Db	XX	
QY	421 YSHSGQKQVLTWPMVPCAGEGPGSMKDGICDKAQNNSACWDGDCSGNSGGSYTAG	Db	XX	

OY	481 GGTTGSIQVGHPMQFGGINSYSVNCNGCANSWLADKFCDQACNVLSGCGDAGDCQDHF 540	XX	The present invention relates to novel human secretory or membrane proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-ADA54071). The coding sequences are useful in the gene therapy of diseases caused by abnormalities of the proteins, e.g. cancer, inflammatory diseases, osteoporosis or neurological disease.
Db	435 -----Fgggnsttysvycnogcanswladkfcdqacnvlscgfdagdcqdhr 480	CC	
OY	541 HELYKVILLPNQHYYIKGECLPYPSFAEVAKRGVEGAYSDNPTRIRHASIANKWTHL 600	CC	
Db	481 HELYKVILLPNQHYYIKGECLPYPSFAEVAKRGVEGAYSDNPTRIRHASIANKWTHL 540	CC	
OY	601 IMMSGMATTIHNLFTONTDEEFKQIIVTVEDEFPKGKLNSTAQGYENLVSPTLIP 660	XX	Sequence 846 AA;
Db	541 IMMSGMATTIHNLFTONTDEEFKQIIVTVEDEFPKGKLNSTAQGYENLVSPTLIP 600	XX	Query Match 91.1%; Score 4468; DB 6; Length 846; .
OY	661 EAEILFEDIPKEKPFKKRHDVNSTRQAEEVKTLVNSLPLDAQSLNTDLQHL 720	CC	Best Local Similarity 99.8%; Pred. No. 0; Mismatches 844; Conservative 1; Indels 0; Gaps 0;
Db	601 EAEILFEDIPKEKPFKKRHDVNSTRQAEEVKTLVNSLPLDAQSLNTDLQHL 660	CC	1 MLFKLILQORTYTCIUSHRIGLYVQFLGWWTTTVAQFQGEVILWSRDQHIVLFDSDYRN 60
OY	721 GDITLGYNLSKALLSPFLMSQHAKTKNQAIITDEINDSIVALPOEKQVHSILPNSLG 780	CC	1 MLFKLILQORTYTCIUSHRIGLYVQFLGWWTTTVAQFQGEVILWSRDQHIVLFDSDYRN 60
Db	661 GDITLGYNLSKALLSPFLMSQHAKTKNQAIITDEINDSIVALPOEKQVHSILPNSLG 720	CC	61 AGKSFQRNLCLPMPIDVVYTWTNGTDIBELKEIQLQVERBOMERBOKANREILGNTTEPK 120
OY	781 VSELRQLITPPAVSVKNGHOGQNPFLDPLDETTARFRVETHQTKIGGNVTEKEPLIV 840	CC	61 AGKSFQRNLCLPMPIDVVYTWTNGTDIBELKEIQLQVERBOMERBOKANREILGNTTEPK 120
Db	721 VSELRQLITPPAVSVKNGHOGQNPFLDPLDETTARFRVETHQTKIGGNVTEKEPLIV 780	CC	121 KSEKQLECLJLTHCKVPMVLVDLAPANTIKOLPLSPFSASDFNVAKPKNPSINV 180
OY	841 PLESQMTEKKITGKEKENSRSMBENABNHIGYEVILGRKLQHYDTSYLPLPWKKYR 900	CC	121 KSEKQLECLJLTHCKVPMVLVDLAPANTIKOLPLSPFSASDFNVAKPKNPSINV 180
Db	781 PLESQMTEKKITGKEKENSRSMBENABNHIGYEVILGRKLQHYDTSYLPLPWKKYR 840	CC	181 SVVVPFDSTKVEDAHSGLKGNRSQTVWGRYLTDKEVPGVLMQDIAFLSGRPPTKET 240
OY	901 QDLDDEESEKLTQLAYFIDSQVNGRQLK 928	CC	181 SVVVPFDSTKVEDAHSGLKGNRSQTVWGRYLTDKEVPGVLMQDIAFLSGRPPTKET 240
Db	841 QDLDDEESEKLTQLAYFIDSQVNGRQLK 868	CC	241 NOLKTLPENLSSKKUQOLYSEASVALKJNPKDQELNKQTKKNTIDSKELTISPA 300
RESULT 10			
ADA54824	ID ADA54824 standard; protein; 846 AA.	Db	241 NOLKTLPENLSSKKUQOLYSEASVALKJNPKDQELNKQTKKNTIDSKELTISPA 300
XX	XX	Db	301 YLWDLDSATISQSKODEDTASRSPEDNEALRYSRISRHAPEWRNTITVNGOIPSMLN 360
AC	ADA54824;	Db	301 YLWDLDSATISQSKODEDTASRSPEDNEALRYSRISRHAPEWRNTITVNGOIPSMLN 360
XX	20-NOV-2003 (first entry)	Db	361 DNPRVTIVTHQDFRNLSHLPLTPSSPAESHTSHIRIESELSQKTYLNDVMFGKDWNDDF 420
DE	Human protein, SEQ ID 2392.	Db	361 DNPRVTIVTHQDFRNLSHLPLTPSSPAESHTSHIRIESELSQKTYLNDVMFGKDWNDDF 420
XX	XX	Db	421 YSSSKGQKYLTMVPPVAGEGCGSWKDGYCDACKNSACWDGGCSGNSGGSRYTAG 480
KW	Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic; Gene Therapy; human; secretory protein; membrane protein; cancer; inflammatory disease; osteoporosis; neurological disease.	Db	421 YSSSKGQKYLTMVPPVAGEGCGSWKDGYCDACKNSACWDGGCSGNSGGSRYTAG 480
OS	Homo sapiens.	Db	481 GGTTGSIQVGHPMQFGGINSYSVNCNGCANSWLADKFCDQACNVLSGCGDAGDCQDHF 540
XX	XX	Db	481 GGTTGSIQVGHPMQFGGINSYSVNCNGCANSWLADKFCDQACNVLSGCGDAGDCQDHF 540
PN	EP1293569-A2.	Db	541 HELYKVILLPNQHYYIKGECLPYPSFAEVAKRGVEGAYSDNPTRIRHASIANKWTHL 600
XX	PD 19-MAR-2-2003.	Db	541 HELYKVILLPNQHYYIKGECLPYPSFAEVAKRGVEGAYSDNPTRIRHASIANKWTHL 600
PA	PA (HELI-) HELIX RES INST. (REAS-) RES ASSOC BIOTECHNOLOGY.	Db	601 IMMSGMATTIHNLFTONTDEEFKQIIVTVEDEFPKGKLNSTAQGYENLVSPTLIP 660
XX	XX	Db	601 IMMSGMATTIHNLFTONTDEEFKQIIVTVEDEFPKGKLNSTAQGYENLVSPTLIP 660
PR	14-SEP-2001; 2001JP-00328381.	Db	661 EAEILFEDIPKEKPFKKRHDVNSTRQAEEVKTLVNSLPLDAQSLNTDLQHL 720
PR	24-JAN-2002; 2002US-0350455P.	Db	661 EAEILFEDIPKEKPFKKRHDVNSTRQAEEVKTLVNSLPLDAQSLNTDLQHL 720
XX	XX	Db	721 GDITLGYNLSKALLSPFLMSQHAKTKNQAIITDEINDSIVALPOEKQVHSILPNSLG 780
WI	WI; 2003-395539/38.	Db	721 GDITLGYNLSKALLSPFLMSQHAKTKNQAIITDEINDSIVALPOEKQVHSILPNSLG 780
DR	N-FSDB; ADA53185.	Db	781 VSELRQLITPPAVSVKNGHOGQNPFLDPLDETTARFRVETHQTKIGGNVTEKEPLIV 840
PT	New polynucleotides encoding full-length polypeptides, e.g. secretory and/or membrane proteins, useful for developing medicines for diseases in which the gene is involved, or as target molecules for gene therapy.	Db	781 VSELRQLITPPAVSVKNGHOGQNPFLDPLDETTARFRVETHQTKIGGNVTEKEPLIV 840
XX	XX	Db	841 PLESQM 846
PS	Claim 14; SEQ ID NO 2392; 205pp; English.	Db	841 PLESQM 846

RESULT 11  
 ABU07381 standard; protein; 1459 AA.  
 XX  
 AC  
 XX  
 DT 28-JAN-2003 (first entry)  
 DE Human protein NOV12.  
 XX  
 KW Human; NOVX; cardiomyopathy; atherosclerosis; cancer; hypertension;  
 KW diabetes; inflammation; autoimmune disorder; allergy; blood disorder;  
 KW acquired immunodeficiency syndrome; AIDS; obesity; asthma;  
 KW immunoglobulin A nephropathy; cirrhosis; arthritis; Alzheimer's disease;  
 KW Parkinson's disease; goitre; infection; stroke; muscular dystrophy;  
 KW epilepsy; wasting disorder; neurogenesis; cell differentiation;  
 KW cell proliferation; haematopoiesis; wound healing; angiogenesis;  
 KW gene therapy; single nucleotide polymorphism; SNP.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200285922-A2.  
 XX  
 PD 31-OCT-2002.  
 XX  
 PR 11-APR-2002; 2002WO-US011634.  
 XX  
 PR 23-APR-2001; 2001US-028574B.  
 PR 24-APR-2001; 2001US-028606B.  
 PR 25-APR-2001; 2001US-028629P.  
 PR 03-MAY-2001; 2001US-028833P.  
 PR 16-MAY-2001; 2001US-0291241P.  
 PR 14-SEP-2001; 2001US-0322284P.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 PA (MILL-) MILLENNIUM PHARM INC.  
 XX  
 PT Pena CEA, Guo X, Shimbetsu RA, Padigaru M, Kekuda R, Spytek KA;  
 PT Meirabani F, Topper JN, Malayanak Um, Wasserman S, Edinger S;  
 PT Smithson G, Gunther E, Komives L;  
 DR WPI; 2003-058712/05.  
 XX  
 PT New NOVX polypeptides and nucleic acids, useful for preventing or  
 PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,  
 PT atherosclerosis or diabetes, and in chromosome mapping, tissue typing or  
 PT pharmacogenomics.  
 XX  
 PS Claim 1; Page 94-95; 301pp; English.

The invention relates to an isolated polypeptide comprising any one of 17 human NOVX (1-9, 10a, 10b, 11-16) appearing as ABU07365-ABU07385, a mature form of it, or a variant of them, where one or more residues of the variant differs in not more than 15 % from the residues of the sequence of them and their encoding poly nucleotides appearing as ABX10223 -ABX10239. Also included are NOVX expression vectors, transformed cells, antibodies, identifying an agent that binds to or modulates the expression or activity of NOVX and screening for a modulator of activity or of latency or predisposition to a NOVX-associated disorder. The NOVX polypeptides, poly nucleotides and antibodies are useful in manufacturing a medicament for treating or preventing a syndrome associated with NOVX-associated disorder, such as cardiomyopathy, atherosclerosis, cancer, hypertension, diabetes, inflammation, autoimmune disorders, allergies, blood disorders, acquired immunodeficiency syndrome (AIDS), obesity, asthma, immunoglobulin (IgA) nephropathy, cirrhosis, arthritis, Alzheimer's disease, Parkinson's disease, goitre, infections (e.g., bacterial, viral, parasitic), stroke, muscular dystrophy, epilepsy, and other wasting disorders associated with chronic diseases. The nucleic acids and polypeptides may also be used as targets for the identification of small molecules that modulate or inhibit e.g. neurogenesis, cell differentiation, cell proliferation, haematopoiesis, wound healing and

CC angio genesis; in gene therapy, in generation of antibodies that bind immunospecifically to NOVX substances for use in therapeutic or diagnostic methods. The nucleic acids are further used as hybridisation probes, in chromosome mapping, tissue typing, preventive medicine, and pharmacogenomics. The polypeptides are also useful as vaccines. The present sequence represents a NOVX protein

XX  
 SQ Sequence 1459 AA;

Query Match 89.5%; Score 4391; DB 6; Length 1459;  
 Best Local Similarity 73.0%; Pred. No. 0;  
 Matches 884; Conservative 13; Mismatches 28; Indels 286; Gaps 13;

Oy 1 MLFKLQLQRQTYTCIHLGHRGLYVCFGLVVUTIVSAFORCE----- 39  
 Db 1 MLFKLQLQRQTYTCIHLGHRGLYVCFGLVVUTIVSAFORCE----- 39  
 Oy 40 ----- 39  
 Db 61 NHPAQVDSAEVKSGIRKGKNGCRAVSLQDWPGTRGCANIFTFAFDCKFSEVSOKRF 120  
 Oy 40 ----- 58  
 Db 121 YILQNCWHLTDWGWTWLMALHGSLLQGPASBPGCVLKAKYVLSRQDYLFLFSDYR 180  
 Oy 59 NIAGKSFRQLCLPMPIDVWVYTWNGDLELIKELQVREQMEEEKAMREILGKTTK 118  
 Db 181 NIAGKSFRQLCLPMPIDVWVYTWNGDLELIKELQVREQMEEEKAMREILGKTTK 240  
 Db 119 TKSSEKQFQECUCLTICKIKUPMLVDPALPANTIKDVSLSYFSFHASDINVAKPKNST 178  
 Db 241 TKSSEKQFQECUCLTICKIKUPMLVDPALPANTIKDVSLSYFSFHASDINVAKPKNST 300  
 Oy 179 NVSVWVFESTKD-----VEDAHSGLKGNS-----ROTW----- 208  
 Db 301 NVSVWVFESTKD-----VEDAHSGLKGNS-----ROTW----- 208  
 Oy 209 -----RGVLTDEKVPGLT-MODLAFLSFGFP 234  
 Db 361 AQJSSALQIITPRVSQRANTSYEIKDTPPLRGA--KEVPGPBTGQPLSFARCLP 417  
 Oy 235 -----PTF-----KET----- 10LKT---KL 247  
 Db 418 TDLRKVNELRDFVWVQMYKQDPSTIHTKETCFLREQEVSMEGESYYKBENTKELKTSKKV 477  
 Oy 248 PEN---LSSKVQLOIS-----EASVALLKANKPKOPEQWQTK 285  
 Db 478 BENISTDELSSEEDSLEIDNEAVIEPTBDPSDPOEMGDEEASYVALLKANKPKOPEQWQTK 537  
 Oy 538 KNTMDIGELTISPAVYLWDGATQSQRQKDEDISARARFEDNEBLRSLRSTERHADWNR 345  
 Db 286 KNTMDIGELTISPAVYLWDGATQSQRQKDEDISARARFEDNEBLRSLRSTERHADWNR 597  
 Oy 346 IFTIVNGQIIPSWLNLNDNPRTVTHQDFNPRNTSHLPFSSPALESIHRTEGLSOPFIYL 405  
 Db 598 IFTIVNGQIIPSWLNLNDNPRTVTHQDFNPRNTSHLPFSSPALESIHRTEGLSOPFIYL 657  
 Oy 405 NDDTMFGKDWPDPYFSISKGORYVLYWPVCAECGPGSMIKDGCDKACNNACWDG 465  
 Db 658 NDDTMFGKDWPDPYFSISKGORYVLYWPVCAECGPGSMIKDGCDKACNNACWDG 717  
 Oy 466 GDGSGNSGSRSYAGGGTAGGGTAGGGTAGGGTAGGGTAGGGTAGGGTAGGGTAGGG 525  
 Db 718 GDGSGNSGSRSYAGGGTAGGGTAGGGTAGGGTAGGGTAGGGTAGGGTAGGGTAGGG 777  
 Oy 526 LSCCPFDAGDCGQ-----DHFHEJYKVLNPQNTYII 557  
 Db 778 LSCCPFDAGDCGQNSDSKRNTRKEEKCPVKKKIMFLPPFLDHFLHELYKVLNPQNTYII 837  
 Oy 558 PKGCLPLPFSAEVAKRGVEGAYSDNPIRHSIANWKWTHLIMSGMNATTIHNLTP 617  
 Db 838 PKGCLPLPFSAEVAKRGVEGAYSDNPIRHSIANWKWTHLIMSGMNATTIHNLTP 897

OY	618 QNTNDERPKMQITVEUDTRGPKUNSTAQKGYENLVSPIITLPEARILFEDIPKRRPK 677	CC patient suffering from a lysosomal storage disease. The present sequence is used in the exemplification of the invention.
Db	898 QNTNDERPKMQITVEUDTRGPKUNSTAQKGYENLVSPIITLPEARILFEDIPKRRPK 957	CC
QY	678 FKRHDVNSTRAQBEVKIPNVLISLUPKDAQLSINTLDOIHEGDTLK3YNLSKALLR 737	XX
Db	958 FKRHDVNSTRAQEEVKIPNVLISLUPKDAQLSINTLDOIHEGDTLK3YNLSKALLR 1017	SQ Sequence 908 AA;
QY	738 SFLMNOSPAKIKKQAITDTEUDSLVAPQEKVKHSLTIPSLVPLSOMKEKKITGEKEK 797	Query Match 79.4%; Score 3894; DB 4; Length 908;
Db	1018 SFLMNOSPAKIKKQAITDTEUDSLVAPQEKVKHSLTIPSLVPLSOMKEKKITGEKEK 1017	Best Local Similarity 79.8%; Pred. No. 3 6e-35; Mismatches 104; Indels 22; Gaps 7; Matches 742; Conservative 61; Mismatches 104; Indels 22; Gaps 7;
QY	798 NGIDOGONPPDLETTARFRVETHTOKIGANVTKKPSLIVPLSOMKEKKITGEKEK 857	Db 1 MLFKLKLQRQTYTCISSLRGGLYVCLGMMVTVAFAQGEWLEWSRQHVLFDSDYRN1 60
Db	1078 NGIDOGONPPDLETTARFRVETHTOKIGANVTKKPSLIVPLSOMKEKKITGEKEK 1077	Db 1 MLFKLKLQRQTYTCISSLRGGLYVCLGMMVTVAFAQGEWLEWSRQHVLFDSDYRN1 60
QY	858 ENSRMEENAENHIGVTEVLGRKLUQYTDPSLIGFLPWEKCYFQDLIDEBESLKIOLAYE 917	QY 61 AGKSFQNRLCLPMIDVYTWNGTDEELKLOQTEQEMERQKAMRSLIGKNTBPTK 120
Db	1138 ENSRMEENAENHIGVTEVLGRKLUQYTDPSLIGFLPWEKCYFQDLIDEBESLKIOLAYE 1197	Db 61 AGKSFQNRLCLPMIDVYTWNGTDEELKLOQTEQEMERQKAMRSLIGKNTBPTK 120
QY	918 TDSKNTERQLK 928	Db 61 AGKSFQNRLCLPMIDVYTWNGTDEELKLOQTEQEMERQKAMRSLIGKNTBPTK 120
Db	1198 TDSKNTERQLK 1208	Db 121 KSEKQDECILTHCTKVPMVLDALPANTLDVPSYPSFASDITFNVAKPNNTNV 180
RESULT 12		Db 121 KSEKQDECILTHCTKVPMVLDALPANTLDVPSYPSFASDITFNVAKPNNTNV 180
ABR61384	ID ABR61384 standard; protein; 908 AA.	Db 181 SVVVFDSKTDKVEDAHSGLIKGNRSOTWGRYLTDKEVPGVLMDQIAFLSGFPPTKET 240
XX	AC ABR61384;	Db 181 PVVVFDTKVEDAHGGFKGGQDIDWNRAYLTDDKAPGLVLIQGIAFLSGFPPTKET 240
XX	DT 01-AUG-2003 (first entry)	Db 241 NOLKTULPEN-LSSKVKIOLSEASVALLKUNNPQFOBLKOTKNTDGKELISP 299
DE	Murine GlcNAc-phosphotransferase alpha-subunit.	Db 241 SOKLKLPRKAFLKIKLURLYEASVALKUNNPQFOBLKOTKNTDGKELISP 300
XX		Db 300 AXILWDLSAISQSKQDPEIASFEDNEELYSLSERHAPMWRNTPVINGQISWLN 359
XX		Db 301 AXILWDLSAISQSKQDPEIASFEDNEELYSLSERHAPMWRNTPVINGQISWLN 360
DB		Db 360 LDNPRVITYTHQDVFRNISHLPTSSPAIESHHRIGLSQKRYIYIADDVMFGKDWWFDD 419
QY		Db 361 LDNPRVITYTHQDIFQNLSHLPTSSPAIESHHRIGLSQKRYIYIADDVMFGKDWWFDD 420
Db		Db 420 FYSSISKOKVYLWTPVPCNAEGCGPSWTKDGYCDKACANNACWDGDCSGNSGGSYIA 479
QY		Db 421 FYSSISKOKVYLWTPVPCNAEGCGPSWTKDGYCDKACANTSFCWDGANCSCGTTAGNRFVA 480
Db		Db 480 GGSGTGTCVGHPWQFGGINSVSYCNGCANSWLAQKFCDOACNVSICGFADGCCDH 539
QY		Db 481 RGGGTGNICAGQHQWQFGGINTSYCNGCANSWLAQKFCDOACNVSICGFADGCCDH 540
Db		Db 540 FHELYKVVILLPQTHYIIPKGCLPYPSFAEVAKRGEGAVSNDPIRHASTANKMTIH 599
QY		Db 541 FHELYKVVILLPQTHYIIPKGCLPYPSFAEVAKRGEGAVSNDPIRHASTANKMTIH 600
Db		Db 600 LIMMSGMATTIHFNTONTDEEPMQITVEUDTRGPKUNSTAQKGYENLVSPIITL 659
QY		Db 601 LIMSGMATTIHFNTONTDEEPMQITVEUDTRGPKUNSTAQKGYENLVSPIITL 660
Db		Db 660 PEAFILFEDIPKERRPKPKRHDVNSTRAQEEVKIPNVLISLUPKDAQLSINTLDOIHEGDTLK3YNLSKALLR 719
QY		Db 661 PQADVPFFEDVPKERPKIRHDVNSTRAQEEVKIPNVLISLUPKDAQLSINTLDOIHEGDTLK3YNLSKALLR 720
XX		Db 720 HGDPTLKGNLKGSALARSPMSQHAKRQDQITBETNDLIVAPQEKVKISPLNLPSL 779
PT		Db 721 RGDPTLKGNLKGSALARSPMSQHAKRQDQITBETNDLIVAPQEKVKISPLNLPSL 777
PT		Db 780 GVSERLORLTFRPVSKYNGHODGONPPDLETTARFRVETHTOKIGANVTKKPSL 839
PT		Db 778 AGEHRSERWTAPATVYGRDHALNPPVLENARI----AQPTGVTVERENISPLI 832
XX		Db 840 VPLESQMKEKKITGEKEGENSMRMEENHIGVTEVLGRKLUQYTDPSLIGFLPWEKCY 899
CC		Db 833 VPPESHLP-----KEEBSDRAGNA---VPKLYPGRRLQ---QNYPGFLPWEKCY 879
CC		QY 900 QDLDDEBSLKIOLAYTDSKRTGRLK 928
CC		Db 880 QDLDDEBSLKIOLAYTDSKRTGRLK 908

The invention relates to a novel isolated human N-acetylglucosamine-1-phosphotransferase (GlcNAc-phosphotransferase) (I) and phosphodiester alpha-GlcNAcCase (N-acetylglucosamine-1-phosphodiester alpha-N-acetylglucosaminidase) (II). The protein of the invention has nephrotropic activity, and may be useful in enzyme replacement therapy. A protein of the invention (I), (II) is useful for preparing a phosphorylated lysosomal hydrolase. The phosphorylated hydrolase comprising a terminal mannose-6-phosphate, is useful for treating a

RESULT 13  
 ADD27817 standard; protein; 908 AA.  
 XX  
 ADD27817; DT 15-JAN-2004 (first entry)  
 XX  
 G1cNAc-phosphotransferase associated protein #1.  
 XX  
 mouse; protein phosphorylation; soluble G1cNAc-phosphotransferase;  
 XX UDP-N-acetylglucosamine; lysosomal storage disease; Fabry's disease; ds.  
 XX  
 Mus musculus.  
 XX  
 PN US2003119088-A1.  
 XX  
 PD 26-JUN-2003.  
 XX  
 PF 21-DEC-2001; 2001US-00023888.  
 XX  
 PR 21-DEC-2001; 2001US-00023888.  
 XX  
 PA (NOVA-) NOVAZYME PHARM INC.  
 XX  
 PI Canfield W, Kudo M;  
 XX  
 DR WPI; 2003-810323/75.  
 XX  
 PT Phosphorylating a protein for treating a patient suffering from a  
 PT lysosomal storage disease e.g. Fabry's disease by contacting the protein  
 PT with a soluble G1cNAc-phosphotransferase and producing a phosphorylated  
 protein.  
 XX  
 Disclosure; SEQ ID NO 9; 55PP; English.  
 XX  
 The invention relates to a method of phosphorylating a protein comprising  
 CC contacting the protein with a soluble G1cNAc-phosphotransferase (UDP-N-  
 CC acetylglucosamine) and producing a phosphorylated protein. The method is  
 CC useful for treating a patient suffering from a lysosomal storage disease  
 e.g. Fabry's disease. The present sequence represents the amino acid  
 sequence of a G1cNAc-phosphotransferase associated protein.  
 XX  
 Sequence 908 AA;  
 SQ

Query Match	79.4%	Score	3894;	DB	7;	Length	908;	
Best Local Similarity	79.9%;	Pred.	No.	3	6e-305;	Mismatches	742;	
Matches	61;	Conservative		104;	Indels	22;	Gaps	7;

QY 1 MLFKLQLQTYTQLSHRYGLYVCFLGIVVTTSAFQFGEVWLEWSQDQHLYLFDSYRNDI 60  
 Do 1 MLFKLQLQTYTQLSHRYGLYVCFLGIVVTTSAFQFGEVWLEWSQDQHLYLFDSYRNDI 60  
 QY 61 AGSFQFLNCLMPIDVYTWNGDLEELKELOQREQMFEEQKAMRLIGRNTTEPK 120  
 Do 61 AGSFQFLNCLMPIDVYTWNGDLEELKELOQREQMFEEQKAMRLIGRNTTEPK 120  
 QY 121 KSFQKOLELLTRCKVPMVLDPALPANTLKDVPSIMPSFASDIFNTVAKPKPSTNV 180  
 Do 121 KSFQKOLELLTRCKVPMVLDPALPANTLKDVPSIMPSFASDIFNTVAKPKPSTNV 180  
 QY 181 SVVVFDFSTDVKEDAHSGJLKGNSRQTVWRGYLTDXEVPGLYTMQDLSAFGSPPTFKEK 240  
 Do 181 SVVVFDFSTDVKEDAHSGJLKGNSRQTVWRGYLTDXEVPGLYTMQDLSAFGSPPTFKEK 240  
 Do 181 PVVVFDTKDVDAHAGFPGKQGQDQTWRAVITTDKDPAGVLIQGIAFLSGPPTFKEK 240  
 QY 241 NQKTKLKEN-1SSKVKLQLYSEASVALLKLNPKDFOELNKOTKRNMTDGKLTISP 299  
 Do 241 NQKTKLKEN-1SSKVKLQLYSEASVALLKLNPKDFOELNKOTKRNMTDGKLTISP 299  
 QY 300 AYLIWDSLASISKDDEDISAERFEDHELYRSRSLERHAPWWRIFLVINGQIPSMW 359  
 Do 301 AYLIWDSLASISKDDEDISAERFEDHELYRSRSLERHAPWWRIFLVINGQIPSMW 360

QY 360 LDNPRTVTIVTHQDVFRNLSHLPTFSPAIESHTHRICLSQKFIYANDDVMEGKDWWPD 419  
 Do 361 LDNPRVTIVTHQDIFONLSHLPTFSPAIESHTHRICLSQKFIYANDDVMEGKDWWPD 420  
 QY 420 FYSISKGQKVYLTTWPPVNCAGGGPSWIKDGYCDAKCNNSACWDGDCSGNISGGGRVIA 479  
 Do 421 FYSISKGQKVYLTTWPPVNCAGGGPSWIKDGYCDAKCNNSACWDGDCSGNISGGGRVIA 480  
 QY 480 GGGGTGTSICVGHPMQPGGEGINSVSYCNGCANSWLAZKFCDOACNVCSCGFDAGDGQDH 539  
 Do 481 RGGGTGNNIGAQOHWQPGGAGINTSYCNGCANSWLAZKFCDOACNVCSCGFDAGDGQDH 540  
 QY 540 FHEIYKVILLPNQHYYIPKGBCBLPYPSFAEYAKRGYEGASDNPIRHASTANKRITH 599  
 Do 541 FHEIYKVILLPNQHYYIPKGBCBLPYPSFAEYAKRGYEGASDNPIRHASTANKRITH 600  
 QY 600 LIMISGMNATTIHNLTQNTDNEFQMITEVDREGPKNSTAQKGYENLVSPTL 659  
 Do 601 LIMPQGMNATTIYNLTQNTDNEFQMITEVDREGPKNSTAQKGYENLVSPTL 660  
 QY 660 PEARLFEDIPEKSKPFKPREKHVNSTRARRQEVKPLVNLSSLPKDAQSLNTLDQLE 719  
 Do 661 PQADVPFDVPKSKPFKIRRDVNTGFRFOEVKPLVNLSSLPKDAQSLNTLDQLE 720  
 QY 720 HDITLKGYNLSKALLSFIMNSQHAKIKNQAIIDTENDSIVAPQEQKPKSILPNSL 779  
 Do 721 RDITLKGYNLSKALLSFIMNSQHAKIKNQAIIDTENDSIVAPQEQKPKSILPNSL 777  
 QY 780 GVSRVLQRUTFPAVSVKUNGHDQGQNPDLDETRARFVETHTOKTIGNYTKEKPSL 839  
 Do 778 AGEHRSERWTAPETVTYKGRHALNPPVLTNRL---AQPIPLGVTYSKENSPLI 832  
 QY 840 VPLUSOMTKEKKITGKECKNSRNEENAHNGUTEVNLGRKLUHYDTSYLGFLPWKKY 899  
 Do 833 VPPESHLP-----KEBEDRAGNA--VPUKELVGRRLQ--QVYPGFLPWKKY 879  
 QY 900 FQDIDEBESLKQLAYDSNTGRQIK 928  
 Do 880 FQDIDEBESLKQLAYDSNTGRQIK 908

RESULT 14  
 ID ABW01491  
 ID ABW01491 standard; protein; 908 AA.  
 XX  
 AC ABW01491;  
 XX  
 DT 15-JAN-2004 (first entry)  
 XX  
 DE Mouse protein #1 used to illustrate the method of the invention.  
 XX  
 Mamrose glycoprotein; gene therapy; carbohydrate deficient cell;  
 KW lysosomal storage disease; gastrointestinal; mouse;  
 KW lectin resistant cell; deoxymannojirimycin; kifunensine;  
 KW glycosylation inhibition.  
 XX  
 Mus musculus.  
 XX  
 PN US2003124652-A1.  
 XX  
 PD 03-JUL-2003.  
 XX  
 PR 21-DEC-2001; 2001US-00023889.  
 XX  
 PR 21-DEC-2001; 2001US-00023889.  
 XX  
 PA (NOVA-) NOVAZYME PHARM INC.  
 XX  
 PI Canfield WM;  
 XX  
 DR WPI; 2003-810984/76.

PT producing a high mannose glycoprotein for treating lysosomal storage  
 PT disease; comprises culturing the lectin resistant mammalian cell in the  
 PT presence of deoxymannojirimycin and kifunenine.

XX Disclosure; Page 23-25; 46pp; English.

The invention relates to a method for producing a high mannose  
 glycoprotein. The method comprises: introducing and expressing a  
 polynucleotide encoding a glycoprotein into a mammalian cell; culturing  
 the cell in the presence of a lectin to obtain a lectin resistant cell;  
 isolating the cell, culturing the cell in the presence of  
 deoxymannojirimycin and kifunenine to inhibit glycosylation of the  
 glycoprotein; and collecting the glycoprotein. The invention is useful in  
 gene therapy. The method is useful for producing a high mannose  
 glycoprotein in a complex carbohydrate deficient cell for treating  
 lysosomal storage disease. The present sequence is mouse protein used to  
 illustrate the method of the invention

Sequence 908 AA;

SQ

Best Local Similarity

Score

DB

Length

908;

Matches

742;

Conservative

Pred. No.

3. 6e-305;

Indels

104;

Gaps

7;

QY

1 MFLKLLQROTCTCLSHRGYLYCPLGVVTTTSAFORGEVLEWSRQHVLFDSDYDNTI 60

Db

1 MFLKLLQROTCTCLSHRGYLYCPLGVVTTTSAFORGEVLEWSRQHVLFDSDYDNTI 60

QY

61 AGKSFONRLCLMPIDVVYTWTNGTDLILKELOQVREHMEBOKARETLGKNTBPTK 120

Db

61 AGKSFONRLCLMPIDVVYTWTNGTDLILKELOQVREHMEBOKARETLGKNTBPTK 120

QY

651 PQLVPPSVPKERRQPKRRAVDNATORQBEVK-PRVNISLIPERQQLNLQI 720

Db

720 HGDITLKSYNLKSALLRSFLMSQHAKIKNGAII-DETNDSLVAPOEQKOVHKSLIPNSI 779

RESULT 15

ID

ABW01540

XX

ABW01540 standard; protein; 908 AA.

AC

ABW01540;

XX

DT

15-JAN-2004 (first entry)

DE

Mouse protein #1 used to illustrate the method of the invention.

XX

KW

Glycoprotein; lectin; lysosomal storage disease; gastrointestinal;

XX

NAcetylglucosamine-1-phosphotransferase; gene therapy; mouse.

OS

Mus musculus.

XX

PN

US2003124653-A1.

XX

PD

03-JUL-2003.

XX

PP

21-DEC-2001; 2001US-00023890.

XX

PR

21-DEC-2001; 2001US-00023890.

XX

PA

(NOVA-) NOVAYME PHARM INC.

XX

PI

Canfield WM;

XX

DR

WPI; 2003-810985/76.

XX

PT

Producing a glycoprotein with reduced complex carbohydrates by culturing

PT

the lectin resistant mammalian cell expressing the glycoprotein for

XX

treating lysosomal storage disease.

XX

PS

Disclosure; Page 23-25; 46pp; English.

XX

CC

The present invention provides a method of producing a glycoprotein

CC

having reduced complex carbohydrates by culturing the lectin resistant

CC

mammalian cell expressing the glycoprotein. The method is useful for

CC

producing a glycoprotein with reduced complex carbohydrates for treating

CC

lysosomal storage disease. The present invention is also useful in gene

CC

therapy. The present sequence is mouse protein used to illustrate the

method of the invention

XX

Sequence 908 AA;

SQ

Best Local Similarity

Score

DB

Length

908;

Matches

742;

Conservative

Pred. No.

3. 6e-305;

Indels

104;

Gaps

7;

QY

1 MFLKLLQROTCTCLSHRGYLYCPLGVVTTTSAFORGEVLEWSRQHVLFDSDYDNTI 60

Db

1 MFLKLLQROTCTCLSHRGYLYCPLGVVTTTSAFORGEVLEWSRQHVLFDSDYDNTI 60

QY

61 AGKSFONRLCLMPIDVVYTWTNGTDLILKELOQVREHMEBOKARETLGKNTBPTK 120

Db

61 AGKSFONRLCLMPIDVVYTWTNGTDLILKELOQVREHMEBOKARETLGKNTBPTK 120

QY

QY 121 KSEKOLECLILTHC1KUPMVLDPALPANI1K0VPSL1YPSPHASD1F1WAKPKN1STNV 180  
 Db 121 KSEKOLECLILTHC1KUPMVLDPALPANI1K0VPSL1YPSPHASD1F1WAKPKN1STNV 180  
 QY 181 SVVFDSTKDVEDAHSGLKGNSR0TVWRYL1TDKEVPG1VIMODA1FLSGFPPTPKET 240  
 Db 181 SVVFDSTKDVEDAHSGLKGNSR0TVWRYL1TDKEVPG1VIMODA1FLSGFPPTPKET 240  
 QY 241 NOLKTKLPEN-LSSKV1KQLYBASVALKUNPKOELNKQTCOMTDGKELTISP 299  
 Db 241 NOLKTKLPEN-LSSKV1KQLYBASVALKUNPKOELNKQTCOMTDGKELTISP 299  
 QY 300 AYLWDLAS1S0SK0DE1SASRFEDNEBLRYSLSRHTHAPWVRN1F1VINGQ1SWLN 359  
 Db 301 AYLWDLAS1S0SK0DE1SASRFEDNEBLRYSLSRHTHAPWVRN1F1VINGQ1SWLN 360  
 QY 360 LDNPRTV1WHD0YFVN1SHLPPSP1IESH1HR1EGLSCKPTY1ND1WFGKDWP 419  
 Db 361 LDNPRTV1WHD0F1ON1SHLPPSP1AESHTHRLLEG1SOK1Y1ND1WFGKDWP 420  
 QY 420 FYSHSKGQKV1L1WVPVCAEGPGSM1KDGCDKACNNSACDMDGGDCSSNSGGSYIA 479  
 Db 421 FYSHSKGQKV1L1WVPVCAEGPGSM1KDGCDKACNNSACDMDGGDCSSNSGGSYIA 480  
 QY 480 GGGATGSG1VGHFMQFGG1NS1SYCNGCANS1WLA1DKFCDOACN1USCGDAGDQDH 539  
 Db 481 RGGGTGNG1GAGQHMQFGG1NT1SYCNGCANS1WLA1DKFCDOACN1USCGDAGDQDH 540  
 QY 540 FHELYKV1V1L1PNQTH1YV1PKGEV1SYF1P1T1R1H1S1ANKWTH 599  
 Db 541 FHELYKV1V1L1PNQTH1YV1PKGEV1SYF1P1T1R1H1S1ANKWTH 600  
 QY 600 L1M1SGMM1W1T1H1L1T1O1T1D1E1K1Q1O1V1T1E1D1R1P1K1L1N1S1T1L1 659  
 Db 601 L1M1GGGM1M1T1Y1N1L1T1Q1M1N1D1E1K1Q1A1V1E1D1R1P1K1L1N1S1T1L1 660  
 QY 650 PEAR1LF1D1K1P1K1R1P1K1R1D1Q1N1D1E1R1C1E1V1K1P1L1N1S1L1P1D1Q1 719  
 Db 651 P0AD1V1P1F1D1V1K1R1D1N1A1T1R1F1E1V1K1P1L1N1S1L1P1K1A1Q1L1N1D1Q1 720  
 QY 720 HGP1T1K1S1Y1N1L1S1A1R1F1M1S1Q1H1A1K1K1N1Q1I1T1E1N1S1V1A1P1Q1K1H1K1S1L1P1N1S1L1 779  
 Db 721 RGP1T1K1S1Y1N1L1S1A1R1F1M1S1Q1H1A1K1K1N1Q1I1T1E1N1S1V1A1P1Q1K1H1K1S1L1P1N1S1L1 777  
 QY 780 G1VER1LQ1T1P1A1V1K1H1G1D1G1Q1N1F1D1E1T1A1R1V1E1T1O1K1G1N1T1K1P1S1L1 839  
 Db 778 AGHR1S1E1W1T1A1P1E1T1V1K1H1G1D1G1Q1N1F1D1E1T1A1R1V1E1T1O1K1G1N1T1K1P1S1L1 832  
 QY 840 VPL1E1S1M1T1R1K1E1K1K1E1N1H1G1V1E1V1G1R1K1Q1H1Y1D1S1Y1G1F1P1W1E1K1Y 899  
 Db 833 V1P1B1S1H1P1----K1E1S1D1E1G1N1A1--V1Y1K1E1V1G1R1Q1--Q1Y1O1F1P1W1E1K1Y 879  
 QY 900 FQD1L1D1E1S1L1K1Q1L1A1F1T1D1S1K1T1G1R1Q1K1 928  
 Db 880 FQD1L1D1E1S1L1K1Q1L1A1F1T1D1S1K1T1G1R1Q1K1 908

Search completed: July 26, 2004, 11:09:08  
 Job time : 90.607 sec

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2004, 11:06:34 ; Search time 26.1576 Seconds

Title: US-10-657-280-1

Perfect score: 4907

Sequence: 1 MLFKLILQRTYTCILSHRYGL.....SLKTQIAYFTDSKNTRQLK 928

Scoring table: BLOSUM62

Gappen 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625571 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:  
 1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:  
 2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:  
 3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:  
 4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:  
 5: /cgn2\_6/ptodata/2/iaa/PCITUS\_COMB.pep:  
 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:  
 ;

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

**SUMMARIES**

Result No.	Score	Query Match Length	DB ID	Description
1	4907	100.0	928	4 US-09-635-872A-1
2	4907	100.0	928	4 US-09-636-072A-1
3	4907	100.0	928	4 US-09-636-060C-1
4	4907	100.0	928	4 US-09-986-551-1
5	3894	79.4	908	4 US-09-635-872A-15
6	3894	79.4	908	4 US-09-636-077A-15
7	3894	79.4	908	4 US-09-636-060C-15
8	3894	79.4	908	4 US-09-916-552-15
9	503	10.3	113	4 US-09-872A-11
10	503	10.3	113	4 US-09-636-077A-11
11	503	10.3	113	4 US-09-636-060C-11
12	479	10.3	113	4 US-09-916-552-11
13	479	9.8	502	4 US-09-635-872A-13
14	479	9.8	502	4 US-09-636-060C-13
15	479	9.8	502	4 US-09-872A-13
16	479	9.8	502	4 US-09-986-552-13
17	265	5.4	545	4 US-08-936-107A-10
18	255	5.2	545	4 US-08-311-731A-248
19	240	4.9	364	4 US-09-328-352-5382
20	168.5	3.4	1015	1 US-08-537-210A-1
21	168.5	3.4	1015	1 US-09-113-822-1
22	168.5	3.4	2471	1 US-08-105-432-19
23	168.5	3.4	2471	1 US-08-083-590-19
24	168.5	3.4	2471	3 US-08-533-384-19
25	168.5	3.4	2471	4 US-08-899-232-1
26	154.5	3.1	2321	4 US-09-230-652-2
27	152.1	3.1	1964	4 US-09-467-997-1

**ALIGNMENTS**

RESULT 1  
US-09-635-872A-1

```

; Sequence 1, Appli
; Sequence 1, Appli
; Patent No. 6533300
; APPLICANT: CANFIELD, WILLIAM
; INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSMAL HYDROLASES
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSMAL HYDROLASES
; FILE REFERENCE: 193613US0
; CURRENT APPLICATION NUMBER: US/09/635,872A
; CURRENT FILING DATE: 2000-09-10
; PRIORITY FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patentin version 3.0
; SEQ ID NO: 1
; LENGTH: 928
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-635-872A-1

```

Query Match 100.0%; Score 4907; DB 4; Length 928;  
 Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFKLILQRTYTCILSHRYGLLYVCPFLGVVTTSAFQIGEVVLEWSRQYHVLFDSDRN 60  
 Db 1 MLFKLILQRTYTCILSHRYGLLYVCPFLGVVTTSAFQIGEVVLEWSRQYHVLFDSDRN 60

QY 61 AGKSRQRNLCLPMPIDVVYTWTNGTDLSEILKEQVQBRMELEQKARERIIGRNTTPK 120  
 Db 61 AGKSRQRNLCLPMPIDVVYTWTNGTDLSEILKEQVQBRMELEQKARERIIGRNTTPK 120

QY 121 KSEKQKCLLTHCKRPMVLDPALPAITLKOPSYPSFSASDIFENVAKPKNPTV 180  
 Db 121 KSEKQKCLLTHCKRPMVLDPALPAITLKOPSYPSFSASDIFENVAKPKNPTV 180

QY 181 SVVVFDTKVEDAISGLKGNSRQTVWGRYLTDKEYPGVLVMDQIAFLSGPPTKET 240  
 Db 181 SVVVFDTKVEDAHSGLKGNSRQTVWGRYLTDKEYPGVLVMDQIAFLSGPPTKET 240

QY 241 NOLKTRLPENLSKVRLQLYSEAVALKUNPKDFEELNKOTKKNNTIDGKLTSPA 300  
 Db 241 NOLKTRLPENLSKVRLQLYSEAVALKUNPKDFEELNKOTKKNNTIDGKLTSPA 300

QY 301 YLMWDLASQSQKDIBDISASRFEDNEEYLRSRSIERHAPWTNTIFTVNGOLPSWNL 360  
 Db 301 YLMWDLASQSQKDIBDISASRFEDNEEYLRSRSIERHAPWTNTIFTVNGOLPSWNL 360

QY 361 DNPRTVTHQDVRLNLSHLPTSSPAESHIRIEGLSOKFTYLNDVNMFGKDWWPPDF 420

Db 361 DNPRVTIVTHQDVFRNLSHLPTSSPALESIHRIEGSOKFTYLNDDMFGKDVPDDE 420  
 Qy 421 YSHSKGQKYLTPWPNCAEGCOSWINGYCDKACNNACWDGGCGSGNSGSRTAG 480  
 Db 421 YSHSKGQKYLTPWPNCAEGCOSWINGYCDKACNNACWDGGCGSGNSGSRTAG 480  
 Qy 481 GGGTGSIGVGHPMQFGGGINSVSVCNQGANSWLAKECDQACNVLSGGFDAGDCODHF 540  
 Db 481 GGGTGSIGVGHPMQFGGGINSVSVCNQGANSWLAKECDQACNVLSGGFDAGDCODHF 540  
 Qy 541 HELVKVILLPQNQHVIIPKGECIIPFSAEVAKRGVGAYSNDPIRHASIAKWKTHL 600  
 Db 541 HELVKVILLPQNQHVIIPKGECIIPFSAEVAKRGVGAYSNDPIRHASIAKWKTHL 600  
 Qy 601 IMMSGNAITHNLTFTONTDEFRKGQTIVEVDTRGPKLNSTRAOKGYENLSPITLP 660  
 Db 601 IMMSGNAITHNLTFTONTDEFRKGQTIVEVDTRGPKLNSTRAOKGYENLSPITLP 660  
 Qy 661 EAELIFEDIPKERCPKPKRHDVNSTRRAOKGVPLNTSLPKQQLSLNTDQLEH 720  
 Db 661 EAELIFEDIPKERCPKPKRHDVNSTRRAOKGVPLNTSLPKQQLSLNTDQLEH 720  
 Qy 721 GDTIKGYNLSKALLSFLMSQHAKKNAOITBETNDLSVAPEQKQVKSILNSLG 780  
 Db 721 GDTIKGYNLSKALLSFLMSQHAKKNAOITBETNDLSVAPEQKQVKSILNSLG 780  
 Qy 721 GDTIKGYNLSKALLSFLMSQHAKKNAOITBETNDLSVAPEQKQVKSILNSLG 780  
 Db 781 VSRRLQRJTFPAYSVKNGHDOQNPIDLETARFVETHQTKTGGVTEKEPKSLIV 840  
 Qy 841 PLESQMTKEKKTGKEKENSRMEENAENHIGTEVULGRKLUQHYSYLSGLPWEKKYF 900  
 Db 841 PLESQMTKEKKTGKEKENSRMEENAENHIGTEVULGRKLUQHYSYLSGLPWEKKYF 900  
 Qy 901 QDLDDEESLKTOLAYFTDSKNTGRQLX 928  
 Db 901 QDLDDEESLKTOLAYFTDSKNTGRQLX 928

RESULT 2  
 US-09-636-077A-1  
 ; Sequence 1 Application US/09636077A.  
 ; Patent No. 653785  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CANFIELD, WILLIAM  
 ; TITER OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE  
 ; FILE REFERENCE: 155612US0  
 ; CURRENT APPLICATION NUMBER: US/09/636,077A  
 ; CURRENT FILING DATE: 2000-08-10  
 ; PRIOR APPLICATION NUMBER: 60/153,831  
 ; PRIOR FILING DATE: 1999-09-14  
 ; NUMBER OF SEQ ID NOS: 52  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 1  
 ; LENGTH: 928  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens

US-09-636-077A-1

Query Match 100.0%; Score 4907; DB 4; Length 928;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MFKLQIORTTOLSLHVGIVCPLGVWVTVSAQFGEVTLNRDQMLVPSYRDN 60  
 Db 1 MFKLQIORTTOLSLHVGIVCPLGVWVTVSAQFGEVTLNRDQMLVPSYRDN 60  
 Qy 61 AGKSFONRCLPMPDITYWVWNGTDLKELQOQREMEEDKAMRISLGKTEPK 120  
 Db 61 AGKSFONRCLPMPDITYWVWNGTDLKELQOQREMEEDKAMRISLGKTEPK 120  
 Qy 121 KSEKOLECLLTHCIKVPMLVLDPALPANILKDVPLSYSPHSASDIFNVAEKPKNPSTNV 180

Db 121 KSEKOLECLLTHCIKVPMLVLDPALPANILKDVPLSYSPHSASDIFNVAEKPKNPSTNV 180  
 Qy 181 SVVPDFSTKOVEDAISGLIKGNNSQTVWRYLTIDKEVPGLVIMQDIAFLSLGPPTFFET 240  
 Db 181 SVVPDFSTKOVEDAISGLIKGNNSQTVWRYLTIDKEVPGLVIMQDIAFLSLGPPTFFET 240  
 Qy 241 NOLKTKUPENLSKVKLIQLYSEASVALLKLNPKDQEBLNKTKONTIDKEBLTSPA 300  
 Db 241 NOLKTKUPENLSKVKLIQLYSEASVALLKLNPKDQEBLNKTKONTIDKEBLTSPA 300  
 Qy 301 YLMDLISAISSQSKODEDASRFEDNEELRSLSIERHAPWVNIFVINGOIPSWNL 360  
 Db 301 YLMDLISAISSQSKODEDASRFEDNEELRSLSIERHAPWVNIFVINGOIPSWNL 360  
 Qy 361 DNPRVTIVTHQDVNLSHLPTSSPALESIHRIEGSOKFTYLNDDMFGKDVPDDE 420  
 Db 361 DNPRVTIVTHQDVNLSHLPTSSPALESIHRIEGSOKFTYLNDDMFGKDVPDDE 420  
 Qy 421 YSHSKGQKYLTPWPNCAEGCOSWINGYCDKACNNACWDGGCGSGNSGSRTAG 480  
 Db 421 YSHSKGQKYLTPWPNCAEGCOSWINGYCDKACNNACWDGGCGSGNSGSRTAG 480  
 Qy 481 GGGTGSIGVGHPMQFGGGINSVSVCNQGANSWLAKECDQACNVLSGGFDAGDCODHF 540  
 Db 481 GGGTGSIGVGHPMQFGGGINSVSVCNQGANSWLAKECDQACNVLSGGFDAGDCODHF 540  
 Qy 541 HELVKVILLPQNQHVIIPKGECIIPFSAEVAKRGVGAYSNDPIRHASIAKWKTHL 600  
 Db 541 HELVKVILLPQNQHVIIPKGECIIPFSAEVAKRGVGAYSNDPIRHASIAKWKTHL 600  
 Qy 601 IMMSGNAITHNLTFTONTDEFRKGQTIVEVDTRGPKLNSTRAOKGVPLNTSLPKQQLSLNTDQLEH 720  
 Db 601 IMMSGNAITHNLTFTONTDEFRKGQTIVEVDTRGPKLNSTRAOKGVPLNTSLPKQQLSLNTDQLEH 720  
 Qy 721 GDTIKGYNLSKALLSFLMSQHAKKNAOITBETNDLSVAPEQKQVKSILNSLG 780  
 Db 721 GDTIKGYNLSKALLSFLMSQHAKKNAOITBETNDLSVAPEQKQVKSILNSLG 780  
 Qy 781 VSRRLQRJTFPAYSVKNGHDOQNPIDLETARFVETHQTKTGGVTEKEPKSLIV 840  
 Db 781 VSRRLQRJTFPAYSVKNGHDOQNPIDLETARFVETHQTKTGGVTEKEPKSLIV 840  
 Qy 841 PLESQMTKEKKTGKEKENSRMEENAENHIGTEVULGRKLUQHYSYLSGLPWEKKYF 900  
 Db 841 PLESQMTKEKKTGKEKENSRMEENAENHIGTEVULGRKLUQHYSYLSGLPWEKKYF 900  
 Qy 901 QDLDDEESLKTOLAYFTDSKNTGRQLX 928  
 Db 901 QDLDDEESLKTOLAYFTDSKNTGRQLX 928

RESULT 3  
 US-09-636-060C-1  
 ; Sequence 1 Application US/09636060C  
 ; Patent No. 6642038  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CANFIELD, WILLIAM M  
 ; TITER OF INVENTION: GLCNAC PHOSPHOTRANSFERASE OF THE LYSOSOMAL TARGETING PATHWAY  
 ; FILE REFERENCE: 210119US0C00T  
 ; CURRENT APPLICATION NUMBER: US/09/636,060C  
 ; CURRENT FILING DATE: 2000-08-10  
 ; PRIOR APPLICATION NUMBER: 60/153,831  
 ; PRIOR FILING DATE: 1999-09-14  
 ; NUMBER OF SEQ ID NOS: 52  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 1  
 ; LENGTH: 928  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens

US-09-636-060C-1

Query Match 100.0%; Score 4907; DB 4; Length 928;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MFKLQIORTTOLSLHVGIVCPLGVWVTVSAQFGEVTLNRDQMLVPSYRDN 60  
 Db 1 MFKLQIORTTOLSLHVGIVCPLGVWVTVSAQFGEVTLNRDQMLVPSYRDN 60  
 Qy 61 AGKSFONRCLPMPDITYWVWNGTDLKELQOQREMEEDKAMRISLGKTEPK 120  
 Db 61 AGKSFONRCLPMPDITYWVWNGTDLKELQOQREMEEDKAMRISLGKTEPK 120  
 Qy 121 KSEKOLECLLTHCIKVPMLVLDPALPANILKDVPLSYSPHSASDIFNVAEKPKNPSTNV 180

*John Melzer and*

US-09-636-060-C-1

; Sequence 1, Application US/09986552

; Patent No. 6670165

; GENERAL INFORMATION:

; APPLICANT: CAMPFIELD, William

; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES

; FILE REFERENCE: 215089US77D1V

; CURRENT FILING NUMBER: US/09/986,552

; CURRENT FILING DATE: 2001-11-09

; PRIORITY NUMBER: 09/635,872

; PRIORITY NUMBER: 60153,831

RESULT 4  
US-09-986-552-1

RESULT 5  
US-09-635-872A-15  
; Sequence 15, Application US/09635872A  
; Patent No. 6534300  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
; FILE REFERENCE: 195613US0  
; CURRENT APPLICATION NUMBER: US/09/635, 872A  
; CURRENT FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153, 831  
; PRIOR FILING DATE: 1999-09-14  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 15  
; LENGTH: 908  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; US-09-635-872A-15

Query Match  
Best Local Similarity 79.4%; Score 3894; DB 4; Length 908;  
Matches 742; Conservative 91; Mismatches 104; Indels 22; Gaps 7;

QY 1 MLFKLQLQTYTCIPLSHRGLYVCPFLGVNTIVSAFQGEVLEWSRDQYHVLFDSDYRDN 60  
DQ 1 MLFKLQLQTYTCIPLSHRGLYVCPFLGVNTIVSAFQGEVLEWSRDQYHVLFDSDYRDN 60

QY 61 AGKSFQNRLCPLMPIDVVTWNGTDIELKELQOREQMEBEBOKAREIGKNTEPTK 120  
DQ 61 AGKSFQNRLCPLMPIDVVTWNGTDIELKELQOREQMEBEBOKAREIGKNTEPTK 120

QY 121 KSRKOLEBLTHIKVPMVLPALPANTILKDVPSLPSHSASDIFTNPKPNSTNV 180  
DQ 121 KSRKOLEBLTHIKVPMVLPALPANTILKDVPSLPSHSASDIFTNPKPNSTNV 180

QY 181 SVVVFDSKDVEDAHSGNLKGNSRQTMWGRGTYTDPKEVPGVLMODAFLSGFPPTFKET 240  
DQ 181 PVVVFDTTRDVEDAHAGFPKGQQDTPWRAVLTDXDAPGLVLIQAFSLGFPPTFKET 240

QY 241 NOLKTKEPEN-ISSKVKLQLYSEASVALKKNNPDKFOELNKQKAMIDGKELTSP 299  
DQ 241 SOKTKLPRKAKFLKIKLRLYSEASVALKKNNPDKFOELNKQKAMIDGKELTSP 300

QY 300 AYLLWDISAISSQKDEDISASRFEDNEBLRSRTERHAWPWRNITIVNGQPSWN 359  
DQ 301 AYLLWDISAISSQKDEDISASRFEDNEBLRSRTERHAWPWRNITIVNGQPSWN 360

QY 360 LDNPRTVTTDOPVRNTHLPLPFSSPAIESHRIEGLSOKFYLANDDMFGKDWPPD 419  
DQ 361 LDNPRTVTTDOPVRNTHLPLPFSSPAIESHRIEGLSOKFYLANDDMFGKDWPPD 420

QY 420 FYHSKGSKVYLTWPVNCAEGCPGSNIDGCKDCKCNNSCDWGGDCGNSGSRYIA 479  
DQ 421 FYHSKGSKVYLTWPVNCAEGCPGSNIDGCKDCKCNSTPCDWDGGNCGNTAGNRFWA 480

QY 480 GGGGTGSGVGHFWQFGGINSVSCNGCANSWLAFCUQACNVLSCFGDAGDGQDH 539

RESULT 6  
US-09-636-077A-15  
; Sequence 15, Application US/09636077A  
; Patent No. 6537785  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM  
; TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE  
; FILE REFERENCE: 195613US0  
; CURRENT APPLICATION NUMBER: US/09/636, 077A  
; CURRENT FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153, 831  
; PRIOR FILING DATE: 1999-09-14  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 15  
; LENGTH: 908  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; US-09-636-077A-15

Query Match  
Best Local Similarity 79.4%; Score 3894; DB 4; Length 908;  
Matches 742; Conservative 91; Mismatches 104; Indels 22; Gaps 7;

QY 1 MLFKLQLQTYTCIPLSHRGLYVCPFLGVNTIVSAFQGEVLEWSRDQYHVLFDSDYRDN 60  
DQ 1 MLFKLQLQTYTCIPLSHRGLYVCPFLGVNTIVSAFQGEVLEWSRDQYHVLFDSDYRDN 60

QY 61 AGKSFQNRLCPLMPIDVVTWNGTDIELKELQOREQMEBEBOKAREIGKNTEPTK 120  
DQ 61 AGKSFQNRLCPLMPIDVVTWNGTDIELKELQOREQMEBEBOKAREIGKNTEPTK 120

QY 121 KSRKOLEBLTHIKVPMVLPALPANTILKDVPSLPSHSASDIFTNPKPNSTNV 180  
DQ 121 KSRKOLEBLTHIKVPMVLPALPANTILKDVPSLPSHSASDIFTNPKPNSTNV 180

QY 181 SVVVFDSKDVEDAHSGNLKGNSRQTMWGRGTYTDPKEVPGVLMODAFLSGFPPTFKET 240  
DQ 181 PVVVFDTTRDVEDAHAGFPKGQQDTPWRAVLTDXDAPGLVLIQAFSLGFPPTFKET 240

QY 241 NOLKTKEPEN-ISSKVKLQLYSEASVALKKNNPDKFOELNKQKAMIDGKELTSP 299  
DQ 241 SOKTKLPRKAKFLKIKLRLYSEASVALKKNNPDKFOELNKQKAMIDGKELTSP 300

QY 300 AYLLWDISAISSQKDEDISASRFEDNEBLRSRTERHAWPWRNITIVNGQPSWN 359  
DQ 301 AYLLWDISAISSQKDEDISASRFEDNEBLRSRTERHAWPWRNITIVNGQPSWN 360

QY 360 LDNPRTVTTDOPVRNTHLPLPFSSPAIESHRIEGLSOKFYLANDDMFGKDWPPD 419  
DQ 361 LDNPRTVTTDOPVRNTHLPLPFSSPAIESHRIEGLSOKFYLANDDMFGKDWPPD 420

QY 420 FYHSKGSKVYLTWPVNCAEGCPGSNIDGCKDCKCNNSCDWGGDCGNSGSRYIA 479  
DQ 421 FYHSKGSKVYLTWPVNCAEGCPGSNIDGCKDCKCNSTPCDWDGGNCGNTAGNRFWA 480

QY 480 GGGGTGSGVGHFWQFGGINSVSCNGCANSWLAFCUQACNVLSCFGDAGDGQDH 539

RESULT 5  
US-09-635-872A-15  
; Sequence 15, Application US/09635872A  
; Patent No. 6534300  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
; FILE REFERENCE: 195613US0  
; CURRENT APPLICATION NUMBER: US/09/635, 872A  
; CURRENT FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153, 831  
; PRIOR FILING DATE: 1999-09-14  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 15  
; LENGTH: 908  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; US-09-635-872A-15

Query Match  
Best Local Similarity 79.4%; Score 3894; DB 4; Length 908;  
Matches 742; Conservative 91; Mismatches 104; Indels 22; Gaps 7;

QY 1 MLFKLQLQTYTCIPLSHRGLYVCPFLGVNTIVSAFQGEVLEWSRDQYHVLFDSDYRDN 60  
DQ 1 MLFKLQLQTYTCIPLSHRGLYVCPFLGVNTIVSAFQGEVLEWSRDQYHVLFDSDYRDN 60

QY 61 AGKSFQNRLCPLMPIDVVTWNGTDIELKELQOREQMEBEBOKAREIGKNTEPTK 120  
DQ 61 AGKSFQNRLCPLMPIDVVTWNGTDIELKELQOREQMEBEBOKAREIGKNTEPTK 120

QY 121 KSRKOLEBLTHIKVPMVLPALPANTILKDVPSLPSHSASDIFTNPKPNSTNV 180  
DQ 121 KSRKOLEBLTHIKVPMVLPALPANTILKDVPSLPSHSASDIFTNPKPNSTNV 180

QY 181 SVVVFDSKDVEDAHSGNLKGNSRQTMWGRGTYTDPKEVPGVLMODAFLSGFPPTFKET 240  
DQ 181 PVVVFDTTRDVEDAHAGFPKGQQDTPWRAVLTDXDAPGLVLIQAFSLGFPPTFKET 240

QY 241 NOLKTKEPEN-ISSKVKLQLYSEASVALKKNNPDKFOELNKQKAMIDGKELTSP 299  
DQ 241 SOKTKLPRKAKFLKIKLRLYSEASVALKKNNPDKFOELNKQKAMIDGKELTSP 300

QY 300 AYLLWDISAISSQKDEDISASRFEDNEBLRSRTERHAWPWRNITIVNGQPSWN 359  
DQ 301 AYLLWDISAISSQKDEDISASRFEDNEBLRSRTERHAWPWRNITIVNGQPSWN 360

QY 360 LDNPRTVTTDOPVRNTHLPLPFSSPAIESHRIEGLSOKFYLANDDMFGKDWPPD 419  
DQ 361 LDNPRTVTTDOPVRNTHLPLPFSSPAIESHRIEGLSOKFYLANDDMFGKDWPPD 420

QY 420 FYHSKGSKVYLTWPVNCAEGCPGSNIDGCKDCKCNNSCDWGGDCGNSGSRYIA 479  
DQ 421 FYHSKGSKVYLTWPVNCAEGCPGSNIDGCKDCKCNSTPCDWDGGNCGNTAGNRFWA 480

QY 480 GGGGTGSGVGHFWQFGGINSVSCNGCANSWLAFCUQACNVLSCFGDAGDGQDH 539

RESULT 7  
US-09-636-060C-15  
; Sequence 15, Application US/09636060C  
; Patent No. 662038  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM W.  
; TITLE OF INVENTION: GLCNAC PHOSPHOTRANSFERASE OF THE LYOSOMAL TARGETING PATHWAY  
; FILE REFERENCE: 2110119US0CONT  
; CURRENT APPLICATION NUMBER: US/09/636, 060C  
; CURRENT FILING DATE: 2000-08-10  
; PRIORITY APPLICATION NUMBER: 60/1153, 831  
; PRIORITY FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO: 15  
; LENGTH: 908  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; US-09-636-060C-15

Query Match 79.4%; Score 3894; DB 4; Length 908;  
Best Local Similarity 79.9%; Pred. No. 0;  
Matches 742; Conservative 61; Mismatches 104; Indels 22; Gaps 7;

Qy 1 MLFKLQLORQTYTCLSHRYGLVYCFGLGVWTIVSAFORGEVLEWSRDQYHVLFDSYRDNT 60  
Db 61 AGKSFQNCLCLPMPIDVYUVTWNGTDIPLLKELQVTRQEMEROKARRELIGKONTBPK 120  
Db 61 AGKSFQNRLCLPMPIDVYUVTWNGTDIPLLKELQVTRQEMEROKARRELIGKONTBPK 120  
Db 121 KSEKQLECLLTHCKVPMVLNDPALPANITLKDVPSKPSFASADFNVAKPNSTNV 180  
Db 121 KSEKQLECLLSCIKVPMVLNDPALPANITLKDVPSKPSFASADFNVAKPNSTNV 180  
Db 181 SVVVFDSSTDVEDAHSGJLKGNSRQTWRGYLTDKEVPGVLIMQDIAFLSCGFPPTRET 240  
Db 181 PVVVFDT-KVEDAHSAGPKGGQTDWRAVTDKAPGLVLIQAFPLSFPPFRET 240  
Qy 241 NOLKTILLEN-USKVQLQYSEASYVALLKUNPKFOELNKQTKNMFTDGKELTISP 299  
Db 241 SOLTKLPLPKAFLKLQLRSEASYVALLKUNPKFOELNKQTKNMFTDGKELTISP 299  
Qy 300 AYLWDLAISQSKQDEBISARFEDDEBLRVSRSERHAPWVNPIRHASIANKWTIH 359  
Db 300 AYLWDLAISQSKQDEBISARFEDDEBLRVSRSERHAPWVNPIRHASIANKWTIH 359  
Db 301 AYLWDLAISQSKQDEBISARFEDDEBLRVSRSERHAPWVNPIRHASIANKWTIH 600  
Qy 360 LDNPRTVTHODVFRNLSHPTSSPALESHIRIGLSOKPVIYNDWFGKDWPDD 419  
Db 361 LDNPRTVTHODIIFQNLNHLPTSSPALESHIRIGLSOKPVIYNDWFGKDWPDD 420  
Qy 420 FYSHSKGQKVYLTWMPVNCAGEGPSWIKDGYCDKAGNSACWDGCGDSGNSSGSIYIA 479  
Db 421 FYSHSKGQKVYLTWMPVNCAGEGPSWIKDGYCDKAGNSACWDGCGDSGNSSGSIYIA 479  
Qy 480 GGGGTGIGVGHMWFQGGINSVSYCNGCANSWLAKFCDCQACNVISCGFDAGDCDH 539  
Db 481 RGGGTGIGVGHMWFQGGINTISYCNGCANSWLAKFCDCQACNVISCGFDAGDCDH 540  
Qy 540 PHELYKVILLPNOTHYLIPKGECIPLYSFAEVAKGVEGAVSDNPPIRHASANKWTIH 599  
Db 541 PHELYKVILLPNOTHYVVPKEVLYISFSFANJARKRLEGTSNDPPIRHASANKWTIH 600  
Qy 600 LIMMSGMATTHENLTFONTDNEEFPQVTEVDTBGPKUNSTACKGKYENLSPITL 659  
Db 601 LIMMSGMATTHENLTFONTDNEEFPQVTEVDTBGPKUNSTACKGKYENLSPITL 659  
Qy 660 PEARILEDIPKKEPKERKPKRKRHDVNACTRCFQEVKILVNLISLPKAQSLNTLQ 719  
Db 661 PQADVPFEDVPKEKPKRKRHDVNACTRCFQEVKILVNLISLPKAQVLNLDLQ 720  
Qy 720 HDGDTLGYNLSKALRSFLMSQHAKINQALITBDTNSLVAPOEQVKHSILNSL 779  
Db 721 RDGDTLGYNLSKALRSFLMSLDPKQA-RDPETKGKLEVPQENPSSRR--PHGF 777  
Qy 780 GVSERLQLRTFPAVSVKYGHDQGONPLDLETTARFVRETHQTKTIGGNYTKEKPSL 839  
Db 778 AGEHRSERWTAPAEVTWGRDHALNPVPVLETPNARL---AQPTIGVTISKENLSP 839  
Qy 840 VPLESQMTKEKKITGKEKENSRMHEENENHIGTEVILGKLOHYHTSYLGPLPWERKYY 899  
Db 833 VPPESHLP-----KEESEDRAGNA---VPVKELVPGRLQ---QNPYGFPLPWERKYY 879  
Qy 900 FQDLDDEBSLKLQYATDSKNTGRQIK 928  
Db 880 FQDLDDEBSLKLQYATDSKNTGRQIK 908

RESULT 8  
US-09-936-552-15  
; Sequence 15, Application US/09986552  
; Patent No. 6670165  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYOSOMAL HYDROLASES  
; FILE REFERENCE: 211089US77DIV  
; CURRENT APPLICATION NUMBER: US/09/986 552

CURRENT FILING DATE: 2001-11-09  
 PRIORITY APPLICATION NUMBER: 09/635,872  
 PRIORITY FILING DATE: 2000-08-10  
 PRIORITY APPLICATION NUMBER: 60/153,831  
 NUMBER OF SEQ ID NOS: 52  
 SOFTWARE: PatentIn version 3.1  
 SEQ ID NO 15  
 LENGTH: 908  
 TYPE: PRT  
 ORGANISM: Mus musculus  
 US-09-986-552-15

Query Match 79.4%; Score 3894; DB 4; Length 908;  
 Best Local Similarity 79.9%; Pred. No. 0; Mismatches 742; Conservative 61; Indels 22; Gaps 7;  
 Matches 104;

Qy 1 MLFKLQLQTYTCIISHRVGLYICFLGVVTTVAFOFGVVLEWSDQYHVLFDSDRNTI 60  
 Db 1 AGKSPQNRCILPPIDWVTTWNGTDLBLLKLQOQREBQKAMREBLLKQIGKNTTEPTK 120  
 Qy 61 AGKSFONNLCLPRAPIDVYTWNGTDLBLLKLQOQREBQKAMREBLLKQIGKNTTEPTK 120  
 Db 61 AGKSFONNLCLPRAPIDVYTWNGTDLBLLKLQOQREBQKAMREBLLKQIGKNTTEPTK 120

Qy 121 KSEKQBLLETHIKVPMVLDALPAINTKDVPSAIPSPRSASDIFNVAKPRTSTV 180  
 Db 121 KSEKQBLLETHIKVPMVLDALPAINTKDVPSAIPSPRSASDIFNVAKPRTSTV 180

Qy 181 SVVYFDSTDVEDAHSGLKGNSRQTWGRYLTDEKVGVLIMODIAFLSGFPPPTKET 240  
 Db 181 PVVVDITDQVEDAHSGLKGNSRQTWGRYLTDEKVGVLIMODIAFLSGFPPPTKET 240

Qy 241 NOLTKLLEN-LSSKVKLQLYSEASVALKKLNPKDFEBLNKOTKNTMDGKELTISP 299  
 Db 241 SOKTKLKRKAFFKIKLRLSEASVALKKLNPKDFEBLNKOTKNTMDGKELTISP 300

Qy 300 AYLWDLSAISQSKODEDISASRFEDNEELRYSLSIERH 359  
 Db 301 AYLWDLSAISQSKODEDISASRFEDNEELRYSLSIERH 359

Qy 350 LDPPRVTVTHQVRNHSHPFSSPALESHIREGLSKPYIANDDVFCKWPPD 419  
 Db 351 LDPPRVTVTHQVRNHSHPFSSPALESHIREGLSKPYIANDDVFCKWPPD 420

Qy 420 FYSHSKGKQVYLTWVPVNAEGPGPSNIKGCDKA CNSACDWGDGDCSNSGSRYTA 479  
 Db 421 FYSHSKGKQVYLTWVPVNAEGPGPSNIKGCDKA CNSACDWGDGDCSNSGSRYTA 480

Qy 480 GGCGTGSTGIVGHFWQFGGINSVSYCNGCANSLWADKFCDOACNTLSCGGDAGQGDH 539  
 Db 481 RGGGTGNIGAGHQWQFGGINTSYCNGCANSLWADKFCDOACNTLSCGGDAGQGDH 540

Qy 540 PHLYKVILLPNOTHYIIPKGICLPPYSAEVAKRGYEGASDNPTRIRHASIANKWTIH 599  
 Db 541 FHELYKVILLPNOTHYIIPKGICLPPYSAEVAKRGYEGASDNPTRIRHASIANKWTIH 600

Qy 500 LIMHSGMMATHTHNLTFONTDAEFKMQITVEDREGPKNSTAQKGENLSPITL 659  
 Db 601 LIMPGGMNATTYFNLTQNDANEEBKTOIAVEVDTREAPKUNSTTQAYESLSPVTP 660

Qy 650 PEARLDPDKERKPFKFRKHDVNSTRAQEVKPLNLSLIPKDAQSLINTDQIE 719  
 Db 651 POADVPFDVDPKCKRPFKFRKHDVNSTRAQEVKPLNLSLIPKDAQSLINTDQIE 720

Qy 720 HGDITLKGSNLSKSAIIFSIMQSQHAKKNOAIIIDTENDLSLVARQBKOVHSLIPSL 779  
 Db 721 RGDITLKGSNLSKSAIIFSIMQSQHAKKNOAIIIDTENDLSLVARQBKOVHSLIPSL 777

Qy 780 GVSBRLOLUTFPASVVKNGHDOGONTPDLDZETARTRVERHTQKIGANTKEKPSLI 839  
 Db 778 AGEHRSERNTAPETVTVKGRDHALNPBPVLSNARD---AQPRIGTVSKENLSPLI 832

RESULT 9  
 US-09-635-872A-11  
 Sequence 11, Application US/09635872A  
 Patent No. 653400

GENERAL INFORMATION:  
 APPLICANT: CANFIELD, WILLIAM  
 TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYOSOMAL HYDROLASES

FILE REFERENCE: 195131US0  
 CURRENT APPLICATION NUMBER: US/09-635,872A  
 CURRENT FILING DATE: 2000-08-10

PRIORITY APPLICATION NUMBER: 60/153,831  
 PRIORITY FILING DATE: 1999-09-14  
 NUMBER OF SEQ ID NOS: 52  
 SOFTWARE: PatentIn version 3.0

SEQ ID NO 11  
 LENGTH: 113  
 TYPE: PRT  
 ORGANISM: Rattus rattus  
 US-09-635-872A-11

Query Match 10.3%; Score 503; DB 4; Length 113;  
 Best Local Similarity 92.5%; Pred. No. 1-1e-16; Mismatches 99; Conservative 4; Indels 0; Gaps 0;  
 Matches 99;

Qy 233 FPPPKPENTKQKPELNUSSKVKLQLYSEASVALKKLNPKDFEBLNKOTKNTMDG 292  
 Db 1 FPPPKPENTKQKPELNUSSKVKLQLYSEASVALKKLNPKDFEBLNKOTKNTMDG 292

Qy 293 KELTISPAVILWLSAISQSKODEDISASRFEDNEELRYSLSIERH 339  
 Db 61 KELTISPAVILWLSAISQSKODEDISASRFEDNEELRYSLSIERH 339

RESULT 10  
 US-09-636-077A-11  
 Sequence 11, Application US/09636077A  
 Patent No. 6537785

GENERAL INFORMATION:  
 APPLICANT: CANFIELD, WILLIAM  
 TITLE OF INVENTION: METHODS OF TREATING LYOSOMAL STORAGE DISEASE

FILE REFERENCE: 195612US0  
 CURRENT APPLICATION NUMBER: US/09/636,077A  
 CURRENT FILING DATE: 2000-08-10  
 PRIORITY APPLICATION NUMBER: 60/153,831

PRIORITY FILING DATE: 1999-09-14  
 NUMBER OF SEQ ID NOS: 52  
 SOFTWARE: PatentIn version 3.0

SEQ ID NO 11  
 LENGTH: 113  
 TYPE: PRT  
 ORGANISM: Rattus rattus  
 US-09-636-077A-11

```

; ; Patent No. 6534300
; ; GENERAL INFORMATION:
; ; APPLICANT: CANFIELD, WILLIAM
; ; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; ; FILE REFERENCE: 196113US0
; ; CURRENT APPLICATION NUMBER: US/09/635,872A
; ; CURRENT FILING DATE: 2000-08-10
; ; PRIOR APPLICATION NUMBER: 60/153,831
; ; PRIOR FILING DATE: 1999-09-14
; ; NUMBER OF SEQ ID NOS: 52
; ; SOFTWARE: Patentin version 3.0
; ; SEQ ID NO 13
; ; LENGTH: 502
; ; TYPE: PRT
; ; ORGANISM: Rattus rattus
; ; US-09-636-060C-11

Query Match          10.3%;  Score 503;  DB 4;  Length 113;
Best Local Similarity 92.5%;  Pred. No. 1.1e-36;  Mismatches 4;  Indels 0;  Gaps 0;
Matches 99;  Conservative 4;  MisMatches 4;  Indels 0;  Gaps 0;

Qy      233 FPPTKEKNEQKTKLPEENLSSKKLQLYASAVALKLNUKPDQELNKQTKNTIDG 292
Db      1 FPPTKEKNEQKTKLPEENLSSKKLQLYASAVALKLNUKPDQELNKQTKNTIDG 60
Qy      293 KELTISPAVLLNDLSAISQSQKDEDISASREDNEELRSIERSIEH 319
Db      61 KELAISPAVLLNDLSAISQSQKDEDISASREDNEELRSIERSIEH 107

RESULT 12
US-09-986-552-11
; Sequence 11, Application US/09986552
; Patent No. 6670165
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 2150890577D7W
; CURRENT APPLICATION NUMBER: US/09/986,552
; CURRENT FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: 09/635,872
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 11
; LENGTH: 113
; TYPE: PRT
; ORGANISM: Rattus rattus
; US-09-986-552-11

Query Match          9.8%;  Score 479;  DB 4;  Length 502;
Best Local Similarity 55.4%;  Pred. No. 1.8e-33;  Mismatches 37;  Indels 0;  Gaps 0;
Matches 82;  Conservative 29;  MisMatches 37;  Indels 0;  Gaps 0;

Qy      322 RFEDNEELRSIERSIEHAPWNTFVNGOIPSMNLDPRTVTTHQDVPRNLSHLP 381
Db      4 RFDDNEELRSIERSIEHAWHVVYIWTNGOIPSPWMLDSYRVTVPHEVIAPDQOLP 63
Qy      382 TFSSPATSHIRIEGLSQSKFYINDYMGKVWPPDFYHSKQKQYLTMWPVPGACG 441
Db      64 TFSSSALEFLHRIPIKSKRFYLNDDIFLAGPPLYPDLYTEAEGVRYQAWMVPGACD 123
Qy      442 CPGSWIKUGYCDKACNNSACDWDGGDCS 469
Db      124 CPWYIGDGAODRHCTDACPQDGDCS 151

RESULT 13
US-09-635-872A-13
; Sequence 13, Application US/09635872A
; ; RESULT 15
; ; US-09-636-060C-13

```

; Sequence 13, Application US/c9636060C  
; Patent No. 6642038  
; GENERAL INFORMATION:  
; APPLICANT: CAMPFIELD, WILLIAM M  
; TITLE OF INVENTION: GLCNAC PHOSPHOTRANSFERASE OF THE LYSOSMAL TARGETING PATHWAY  
; FILE REFERENCE: 210119US0CONT  
; CURRENT APPLICATION NUMBER: US/09/636,060C  
; CURRENT FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153, 831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 13  
; LENGTH: 502  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-09-636-060C-13

Query Match 9.8%; Score 479; DB 4; Length 502;  
Best Local Similarity 55.4%; Prod. No. 1.8e-33; Mismatches 37; Indels 0; Gaps 0;  
Matches 82; Conservative 29; MisMatches 37; Indels 0; Gaps 0;

QY 322 RFEDNEEIRYLSIERSIERHAPWVNIFTVNGQIPSWMLDNPRVTIVTHQDVFRNLSHLP 381  
Db 4 RFDDKNERIYLSIERSLERKHAMRHWVITVNGQIPSWMLDSTERVTVPHEVLAPPQLP 63

QY 382 TESPATISHIRTEGSKQKFITNDVNGKDWPPDPYHSKGOKVYLWPPVPCAG 441  
Db 64 TFSSSAITFLHRIPKLSSRFKLYNDIFLGPLAPLYPEDLYTEREGVRVYQAWMVPGCALD 123

QY 442 CPWSWIKGCGDRACNNSACDWWGGDS 469

Db 124 CPWRYIGDGACDRHNCIDACQFDGGDCS 151

Search completed: July 26, 2004, 11:13:18  
Job time : 28.1576 secs

GenCore version 5.1.6  
 Copyright (c) 1993 - 2004 Compugen Ltd.

### Om protein - protein search, using sw model

Run on: July 26, 2004, 11:11:44 ; Search time 71.3389 Seconds  
 (without alignments) 4073.614 Million cell updates/sec

Title: US-10-657-280-1

Perfect score: 4907

Sequence: 1 MFLFKLQLQRTQYICLSHRIGLYVQFLGIVNTIVSAFQGEVLEWSRDQYHVLFDSD'RDNT

Scoring table: BLOSUM62

Gapext 10.0 , Gapext 0.5

Searched: 1288442 seqs, 313154207 residues

Total number of hits satisfying chosen parameters: 1288442

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/prodata/1/pupba/us07\_PUBCOMB.pep:\*

2: /cgn2\_6/ptodata/1/pupba/PCT\_NEW\_PUB.pep:\*

3: /cgn2\_6/ptodata/1/pupba/us06\_NEW\_PUB.pep:\*

4: /cgn2\_6/ptodata/1/pupba/us07\_PUBCOMB.pep:\*

5: /cgn2\_6/ptodata/1/pupba/us08\_PUBCOMB.pep:\*

6: /cgn2\_6/ptodata/1/pupba/us09\_PUBCOMB.pep:\*

7: /cgn2\_6/ptodata/1/pupba/us09\_PUBCOMB.pep:\*

8: /cgn2\_6/ptodata/1/pupba/us09\_PUBCOMB.pep:\*

9: /cgn2\_6/ptodata/1/pupba/us09\_PUBCOMB.pep:\*

10: /cgn2\_6/ptodata/1/pupba/us09\_PUBCOMB.pep:\*

11: /cgn2\_6/ptodata/1/pupba/us09\_PUBCOMB.pep:\*

12: /cgn2\_6/ptodata/1/pupba/us09\_NEW\_PUB.pep:\*

13: /cgn2\_6/ptodata/1/pupba/us10\_PUBCOMB.pep:\*

14: /cgn2\_6/ptodata/1/pupba/us10\_PUBCOMB.pep:\*

15: /cgn2\_6/ptodata/1/pupba/us10\_PUBCOMB.pep:\*

16: /cgn2\_6/ptodata/1/pupba/us10\_NEW\_PUB.pep:\*

17: /cgn2\_6/ptodata/1/pupba/us60\_NEW\_PUB.pep:\*

18: /cgn2\_6/ptodata/1/pupba/us60\_PUBCOMB.pep:\*

### RESULT 1

; Sequence 1, Application US/09895072-1

; GENERAL INFORMATION:

; APPLICANT: CAMPFIELD, WILLIAM M

; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES

; FILE REFERENCE: 200119US0CON

; CURRENT APPLICATION NUMBER: US/09/895,072

; CURRENT FILING DATE: 2001-07-02

; PRIORITY FILING DATE: 2000-08-10

; PRIORITY APPLICATION NUMBER: 60/153,831

; PRIORITY FILING DATE: 2000-08-10

; NUMBER OF SEQ ID NOS: 52

; SOFTWARE: Patentin version 3.1

; SEQ ID NO: 1

; LENGTH: 928

; TYPE: PRT

; ORGANISM: Homo sapiens  
 US-09-895-072-1

### ALIGNMENTS

Result No.	Score	Query Length	DB ID	Description
1	4907	100.0	928	9 US-09-895-072-1
2	4907	100.0	928	9 US-09-895-072-1
3	4907	100.0	928	14 US-10-023-888-4
4	4907	100.0	928	14 US-10-023-889-4
5	4907	100.0	928	14 US-10-023-890-4
6	4907	100.0	928	14 US-10-024-197-4
7	4907	100.0	928	14 US-10-023-894-1
8	4907	100.0	928	14 US-10-023-886-4
9	4630	94.4	1199	14 US-10-023-888-2
10	4630	94.4	1199	14 US-10-023-889-2
11	4630	94.4	1199	14 US-10-023-890-2
12	4630	94.4	1199	14 US-10-024-197-2
13	4630	94.4	1199	14 US-10-023-881-2
14	4473	91.2	847	15 US-10-120-801-88
15	4468	91.1	846	15 US-10-094-749-2392

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

### SUMMARIES

Query Match Best Local Similarity 100.0%; Score 4907; DB 9; Length 928;

Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Sequence 1, Appli

Sequence 2, Appli

Sequence 3, Appli

Sequence 4, Appli

Sequence 5, Appli

Sequence 6, Appli

Sequence 7, Appli

Sequence 8, Appli

Sequence 9, Appli

Sequence 10, Appli

Sequence 11, Appli

Sequence 12, Appli

Sequence 13, Appli

Sequence 14, Appli

Sequence 15, Appli

Sequence 16, Appli

Sequence 17, Appli

Sequence 26, Appli  
 Sequence 15, Appli  
 Sequence 14, Appli  
 Sequence 13, Appli  
 Sequence 12, Appli  
 Sequence 11, Appli  
 Sequence 10, Appli  
 Sequence 9, Appli  
 Sequence 8, Appli  
 Sequence 7, Appli  
 Sequence 6, Appli  
 Sequence 5, Appli  
 Sequence 4, Appli  
 Sequence 3, Appli  
 Sequence 2, Appli  
 Sequence 1, Appli

Sequence 26, Appli  
 Sequence 15, Appli  
 Sequence 14, Appli  
 Sequence 13, Appli  
 Sequence 12, Appli  
 Sequence 11, Appli  
 Sequence 10, Appli  
 Sequence 9, Appli  
 Sequence 8, Appli  
 Sequence 7, Appli  
 Sequence 6, Appli  
 Sequence 5, Appli  
 Sequence 4, Appli  
 Sequence 3, Appli  
 Sequence 2, Appli  
 Sequence 1, Appli

Sequence 26, Appli  
 Sequence 15, Appli  
 Sequence 14, Appli  
 Sequence 13, Appli  
 Sequence 12, Appli  
 Sequence 11, Appli  
 Sequence 10, Appli  
 Sequence 9, Appli  
 Sequence 8, Appli  
 Sequence 7, Appli  
 Sequence 6, Appli  
 Sequence 5, Appli  
 Sequence 4, Appli  
 Sequence 3, Appli  
 Sequence 2, Appli  
 Sequence 1, Appli

Sequence 26, Appli  
 Sequence 15, Appli  
 Sequence 14, Appli  
 Sequence 13, Appli  
 Sequence 12, Appli  
 Sequence 11, Appli  
 Sequence 10, Appli  
 Sequence 9, Appli  
 Sequence 8, Appli  
 Sequence 7, Appli  
 Sequence 6, Appli  
 Sequence 5, Appli  
 Sequence 4, Appli  
 Sequence 3, Appli  
 Sequence 2, Appli  
 Sequence 1, Appli

Sequence 26, Appli  
 Sequence 15, Appli  
 Sequence 14, Appli  
 Sequence 13, Appli  
 Sequence 12, Appli  
 Sequence 11, Appli  
 Sequence 10, Appli  
 Sequence 9, Appli  
 Sequence 8, Appli  
 Sequence 7, Appli  
 Sequence 6, Appli  
 Sequence 5, Appli  
 Sequence 4, Appli  
 Sequence 3, Appli  
 Sequence 2, Appli  
 Sequence 1, Appli

Sequence 26, Appli  
 Sequence 15, Appli  
 Sequence 14, Appli  
 Sequence 13, Appli  
 Sequence 12, Appli  
 Sequence 11, Appli  
 Sequence 10, Appli  
 Sequence 9, Appli  
 Sequence 8, Appli  
 Sequence 7, Appli  
 Sequence 6, Appli  
 Sequence 5, Appli  
 Sequence 4, Appli  
 Sequence 3, Appli  
 Sequence 2, Appli  
 Sequence 1, Appli

Sequence 26, Appli  
 Sequence 15, Appli  
 Sequence 14, Appli  
 Sequence 13, Appli  
 Sequence 12, Appli  
 Sequence 11, Appli  
 Sequence 10, Appli  
 Sequence 9, Appli  
 Sequence 8, Appli  
 Sequence 7, Appli  
 Sequence 6, Appli  
 Sequence 5, Appli  
 Sequence 4, Appli  
 Sequence 3, Appli  
 Sequence 2, Appli  
 Sequence 1, Appli

Sequence 26, Appli  
 Sequence 15, Appli  
 Sequence 14, Appli  
 Sequence 13, Appli  
 Sequence 12, Appli  
 Sequence 11, Appli  
 Sequence 10, Appli  
 Sequence 9, Appli  
 Sequence 8, Appli  
 Sequence 7, Appli  
 Sequence 6, Appli  
 Sequence 5, Appli  
 Sequence 4, Appli  
 Sequence 3, Appli  
 Sequence 2, Appli  
 Sequence 1, Appli

Sequence 26, Appli  
 Sequence 15, Appli  
 Sequence 14, Appli  
 Sequence 13, Appli  
 Sequence 12, Appli  
 Sequence 11, Appli  
 Sequence 10, Appli  
 Sequence 9, Appli  
 Sequence 8, Appli  
 Sequence 7, Appli  
 Sequence 6, Appli  
 Sequence 5, Appli  
 Sequence 4, Appli  
 Sequence 3, Appli  
 Sequence 2, Appli  
 Sequence 1, Appli

Sequence 26, Appli  
 Sequence 15, Appli  
 Sequence 14, Appli  
 Sequence 13, Appli  
 Sequence 12, Appli  
 Sequence 11, Appli  
 Sequence 10, Appli  
 Sequence 9, Appli  
 Sequence 8, Appli  
 Sequence 7, Appli  
 Sequence 6, Appli  
 Sequence 5, Appli  
 Sequence 4, Appli  
 Sequence 3, Appli  
 Sequence 2, Appli  
 Sequence 1, Appli

Sequence 26, Appli  
 Sequence 15, Appli  
 Sequence 14, Appli  
 Sequence 13, Appli  
 Sequence 12, Appli  
 Sequence 11, Appli  
 Sequence 10, Appli  
 Sequence 9, Appli  
 Sequence 8, Appli  
 Sequence 7, Appli  
 Sequence 6, Appli  
 Sequence 5, Appli  
 Sequence 4, Appli  
 Sequence 3, Appli  
 Sequence 2, Appli  
 Sequence 1, Appli

Sequence 26, Appli  
 Sequence 15, Appli  
 Sequence 14, Appli  
 Sequence 13, Appli  
 Sequence 12, Appli  
 Sequence 11, Appli  
 Sequence 10, Appli  
 Sequence 9, Appli  
 Sequence 8, Appli  
 Sequence 7, Appli  
 Sequence 6, Appli  
 Sequence 5, Appli  
 Sequence 4, Appli  
 Sequence 3, Appli  
 Sequence 2, Appli  
 Sequence 1, Appli

Sequence 26, Appli  
 Sequence 15, Appli  
 Sequence 14, Appli  
 Sequence 13, Appli  
 Sequence 12, Appli  
 Sequence 11, Appli  
 Sequence 10, Appli  
 Sequence 9, Appli  
 Sequence 8, Appli  
 Sequence 7, Appli  
 Sequence 6, Appli  
 Sequence 5, Appli  
 Sequence 4, Appli  
 Sequence 3, Appli  
 Sequence 2, Appli  
 Sequence 1, Appli

Sequence 26, Appli  
 Sequence 15, Appli  
 Sequence 14, Appli  
 Sequence 13, Appli  
 Sequence 12, Appli  
 Sequence 11, Appli  
 Sequence 10, Appli  
 Sequence 9, Appli  
 Sequence 8, Appli  
 Sequence 7, Appli  
 Sequence 6, Appli  
 Sequence 5, Appli  
 Sequence 4, Appli  
 Sequence 3, Appli  
 Sequence 2, Appli  
 Sequence 1, Appli

Sequence 26, Appli  
 Sequence 15, Appli  
 Sequence 14, Appli  
 Sequence 13, Appli  
 Sequence 12, Appli  
 Sequence 11, Appli  
 Sequence 10, Appli  
 Sequence 9, Appli  
 Sequence 8, Appli  
 Sequence 7, Appli  
 Sequence 6, Appli  
 Sequence 5, Appli  
 Sequence 4, Appli  
 Sequence 3, Appli  
 Sequence 2, Appli  
 Sequence 1, Appli

Sequence 26, Appli  
 Sequence 15, Appli  
 Sequence 14, Appli  
 Sequence 13, Appli  
 Sequence 12, Appli  
 Sequence 11, Appli  
 Sequence 10, Appli  
 Sequence 9, Appli  
 Sequence 8, Appli  
 Sequence 7, Appli  
 Sequence 6, Appli  
 Sequence 5, Appli  
 Sequence 4, Appli  
 Sequence 3, Appli  
 Sequence 2, Appli  
 Sequence 1, Appli

Sequence 26, Appli  
 Sequence 15, Appli  
 Sequence 14, Appli  
 Sequence 13, Appli  
 Sequence 12, Appli  
 Sequence 11, Appli  
 Sequence 10, Appli  
 Sequence 9, Appli  
 Sequence 8, Appli  
 Sequence 7, Appli  
 Sequence 6, Appli  
 Sequence 5, Appli  
 Sequence 4, Appli  
 Sequence 3, Appli  
 Sequence 2, Appli  
 Sequence 1, Appli

Sequence 26, Appli  
 Sequence 15, Appli  
 Sequence 14, Appli  
 Sequence 13, Appli  
 Sequence 12, Appli  
 Sequence 11, Appli  
 Sequence 10, Appli  
 Sequence 9, Appli  
 Sequence 8, Appli  
 Sequence 7, Appli  
 Sequence 6, Appli  
 Sequence 5, Appli  
 Sequence 4, Appli  
 Sequence 3, Appli  
 Sequence 2, Appli  
 Sequence 1, Appli

Sequence 26, Appli  
 Sequence 15, Appli  
 Sequence 14, Appli  
 Sequence 13, Appli  
 Sequence 12, Appli  
 Sequence 11, Appli  
 Sequence 10, Appli  
 Sequence 9, Appli  
 Sequence 8, Appli  
 Sequence 7, Appli  
 Sequence 6, Appli  
 Sequence 5, Appli  
 Sequence 4, Appli  
 Sequence 3, Appli  
 Sequence 2, Appli  
 Sequence 1, Appli

Sequence 26, Appli  
 Sequence 15, Appli  
 Sequence 14, Appli  
 Sequence 13, Appli  
 Sequence 12, Appli  
 Sequence 11, Appli  
 Sequence 10, Appli  
 Sequence 9, Appli  
 Sequence 8, Appli  
 Sequence 7, Appli  
 Sequence 6, Appli  
 Sequence 5, Appli  
 Sequence 4, Appli  
 Sequence 3, Appli  
 Sequence 2, Appli  
 Sequence 1, Appli

Sequence 26, Appli  
 Sequence 15, Appli  
 Sequence 14, Appli  
 Sequence 13, Appli  
 Sequence 12, Appli  
 Sequence 11, Appli  
 Sequence 10, Appli  
 Sequence 9, Appli  
 Sequence 8, Appli  
 Sequence 7, Appli  
 Sequence 6, Appli  
 Sequence 5, Appli  
 Sequence 4, Appli  
 Sequence 3, Appli  
 Sequence 2, Appli  
 Sequence 1, Appli

Sequence 26, Appli  
 Sequence 15, Appli  
 Sequence 14, Appli  
 Sequence 13, Appli  
 Sequence 12, Appli  
 Sequence 11, Appli  
 Sequence 10, Appli  
 Sequence 9, Appli  
 Sequence 8, Appli  
 Sequence 7, Appli  
 Sequence 6, Appli  
 Sequence 5, Appli  
 Sequence 4, Appli  
 Sequence 3, Appli  
 Sequence 2, Appli  
 Sequence 1, Appli

Sequence 26, Appli  
 Sequence 15, Appli  
 Sequence 14, Appli  
 Sequence 13, Appli  
 Sequence 12, Appli  
 Sequence 11, Appli  
 Sequence 10, Appli  
 Sequence 9, Appli  
 Sequence 8, Appli  
 Sequence 7, Appli  
 Sequence 6, Appli  
 Sequence 5, Appli  
 Sequence 4, Appli  
 Sequence 3, Appli  
 Sequence 2, Appli  
 Sequence 1, Appli

Sequence 26, Appli  
 Sequence 15, Appli  
 Sequence 14, Appli  
 Sequence 13, Appli  
 Sequence 12, Appli  
 Sequence 11, Appli  
 Sequence 10, Appli  
 Sequence 9, Appli  
 Sequence 8, Appli  
 Sequence 7, Appli  
 Sequence 6, Appli  
 Sequence 5, Appli  
 Sequence 4, Appli  
 Sequence 3, Appli  
 Sequence 2, Appli  
 Sequence 1, Appli

Sequence 26, Appli  
 Sequence 15, Appli  
 Sequence 14, Appli  
 Sequence 13, Appli  
 Sequence 12, Appli  
 Sequence 11, Appli  
 Sequence 10, Appli  
 Sequence 9, Appli  
 Sequence 8, Appli  
 Sequence 7, Appli  
 Sequence 6, Appli  
 Sequence 5, Appli  
 Sequence 4, Appli  
 Sequence 3, Appli  
 Sequence 2, Appli  
 Sequence 1, Appli

Sequence 26, Appli  
 Sequence 15, Appli  
 Sequence 14, Appli  
 Sequence 13, Appli  
 Sequence 12, Appli  
 Sequence 11, Appli  
 Sequence 10, Appli  
 Sequence 9, Appli  
 Sequence 8, Appli  
 Sequence 7, Appli  
 Sequence 6, Appli  
 Sequence 5, Appli  
 Sequence 4, Appli  
 Sequence 3, Appli  
 Sequence 2, Appli  
 Sequence 1, Appli

Sequence 26, Appli  
 Sequence 15, Appli  
 Sequence 14, Appli  
 Sequence 13, Appli  
 Sequence 12, Appli  
 Sequence 11, Appli  
 Sequence 10, Appli  
 Sequence 9, Appli  
 Sequence 8, Appli  
 Sequence 7, Appli  
 Sequence 6, Appli  
 Sequence 5, Appli  
 Sequence 4, Appli  
 Sequence 3, Appli  
 Sequence 2, Appli  
 Sequence 1, Appli

Sequence 26, Appli  
 Sequence 15, Appli  
 Sequence 14, Appli  
 Sequence 13, Appli  
 Sequence 12, Appli  
 Sequence 11, Appli  
 Sequence 10, Appli  
 Sequence 9, Appli  
 Sequence 8, Appli  
 Sequence 7, Appli  
 Sequence 6, Appli  
 Sequence 5, Appli  
 Sequence 4, Appli  
 Sequence 3, Appli  
 Sequence 2, Appli  
 Sequence 1, Appli

Sequence 26, Appli  
 Sequence 15, Appli  
 Sequence 14, Appli  
 Sequence 13, Appli  
 Sequence 12, Appli  
 Sequence 11, Appli  
 Sequence 10, Appli  
 Sequence 9, Appli  
 Sequence 8, Appli  
 Sequence 7, Appli  
 Sequence 6, Appli  
 Sequence 5, Appli  
 Sequence 4, Appli  
 Sequence 3, Appli  
 Sequence 2, Appli  
 Sequence 1, Appli

Sequence 26, Appli  
 Sequence 15, Appli  
 Sequence 14, Appli  
 Sequence 13, Appli  
 Sequence 12, Appli  
 Sequence 11, Appli  
 Sequence 10, Appli  
 Sequence 9, Appli  
 Sequence 8, Appli  
 Sequence 7, Appli  
 Sequence 6, Appli  
 Sequence 5, Appli  
 Sequence 4, Appli  
 Sequence 3, Appli  
 Sequence 2, Appli  
 Sequence 1, Appli

Sequence 26, Appli  
 Sequence 15, Appli  
 Sequence 14, Appli  
 Sequence 13, Appli  
 Sequence 12, Appli  
 Sequence 11, Appli  
 Sequence 10, Appli  
 Sequence 9, Appli  
 Sequence 8, Appli  
 Sequence 7, Appli  
 Sequence 6, Appli  
 Sequence 5, Appli  
 Sequence 4, Appli  
 Sequence 3, Appli  
 Sequence 2, Appli  
 Sequence 1, Appli

Sequence 26, Appli  
 Sequence 15, Appli  
 Sequence 14, Appli  
 Sequence 13, Appli  
 Sequence 12, Appli  
 Sequence 11, Appli  
 Sequence 10, Appli  
 Sequence 9, Appli  
 Sequence 8, Appli  
 Sequence 7, Appli  
 Sequence 6, Appli  
 Sequence 5, Appli  
 Sequence 4, Appli  
 Sequence 3, Appli  
 Sequence 2, Appli  
 Sequence 1, Appli

Sequence 26, Appli  
 Sequence 15, Appli  
 Sequence 14, Appli  
 Sequence 13, Appli  
 Sequence 12, Appli  
 Sequence 11, Appli  
 Sequence 10, Appli  
 Sequence 9, Appli  
 Sequence 8, Appli  
 Sequence 7, Appli  
 Sequence 6, Appli  
 Sequence 5, Appli  
 Sequence 4, Appli  
 Sequence 3, Appli  
 Sequence 2, Appli  
 Sequence 1, Appli

Sequence 26, Appli  
 Sequence 15, Appli  
 Sequence 14, Appli  
 Sequence 13, Appli  
 Sequence 12, Appli  
 Sequence 11, Appli  
 Sequence 10, Appli  
 Sequence 9, Appli  
 Sequence 8, Appli  
 Sequence 7, Appli  
 Sequence 6, Appli  
 Sequence 5, Appli  
 Sequence 4, Appli  
 Sequence 3, Appli  
 Sequence 2, Appli  
 Sequence 1, Appli

Sequence 26, Appli  
 Sequence 15, Appli  
 Sequence 14, Appli  
 Sequence 13, Appli  
 Sequence 12, Appli  
 Sequence 11, Appli  
 Sequence 10, Appli  
 Sequence 9, Appli  
 Sequence 8, Appli  
 Sequence 7, Appli  
 Sequence 6, Appli  
 Sequence 5, Appli  
 Sequence 4, Appli  
 Sequence 3, Appli  
 Sequence 2, Appli  
 Sequence 1, Appli

Sequence 26, Appli  
 Sequence 15, Appli  
 Sequence 14, Appli  
 Sequence 13, Appli  
 Sequence 12, Appli  
 Sequence 11, Appli  
 Sequence 10, Appli  
 Sequence 9, Appli  
 Sequence 8, Appli  
 Sequence 7, Appli<br

QY 181 SVVVFDSKVEDAHSGLKGNSRQTWRYGLTIDKEVPGVLMODIAFLSGFRPPFKET 240

DB 181 SVVVFDSKVEDAHSGLKGNSRQTWRYGLTIDKEVPGVLMODIAFLSGFRPPFKET 240

QY 241 NOLKTKLBNLSSKVKULQYSEASVALKLNPQFOELNKOTKONTIDGKELTSPA 300

DB 241 NOLKTKLBNLSSKVKULQYSEASVALKLNPQFOELNKOTKONTIDGKELTSPA 300

QY 301 YLWDLISLTSQSQDEDIISASFEDNEELRSYLSRSERHAPVNIFIVINGQISWLNL 360

DB 301 YLWDLISLTSQSQDEDIISASFEDNEELRSYLSRSERHAPVNIFIVINGQISWLNL 360

QY 351 DNRPVTITHQDFRNLSHLPSSPALESHIRIGLQSQFYIYLDDVMFGKDWPDD 420

DB 351 DNRPVTITHQDFRNLSHLPSSPALESHIRIGLQSQFYIYLDDVMFGKDWPDD 420

QY 421 YSHSKGQKVLTWVPVNAEGPGSWIKDGYDKACNSACWDGGDCSGNSGGSRYTAG 480

DB 421 YSHSKGQKVLTWVPVNAEGPGSWIKDGYDKACNSACWDGGDCSGNSGGSRYTAG 480

QY 481 GGGTGSISGVGHPMQFGGGINSVSYCNGCANSLWAKFCDCQACNVLSGFGDQDHF 540

DB 481 GGGTGSISGVGHPMQFGGGINSVSYCNGCANSLWAKFCDCQACNVLSGFGDQDHF 540

QY 541 HEYKVILLPNQHYYIPKGECIPLYSFAEVAKRGVEGAYDNPIRHASTANKWTHL 600

DB 541 HEYKVILLPNQHYYIPKGECIPLYSFAEVAKRGVEGAYDNPIRHASTANKWTHL 600

QY 601 IMMSGMATTIHLNTFONTDEEKFQKITVEVDTRGPUNKNSTAQKGYENLVSPTLLP 660

DB 601 IMMSGMATTIHLNTFONTDEEKFQKITVEVDTRGPUNKNSTAQKGYENLVSPTLLP 660

QY 651 EARTLFEPKKEGRPKERHVNSTRRAQEVKIPLNISLPLQALSLNTDQLEH 720

DB 651 EARTLFEPKKEGRPKERHVNSTRRAQEVKIPLNISLPLQALSLNTDQLEH 720

QY 721 GDTIKGNLNSKALLSFLMSQHAKIKNQAITBTDNSVAVPOEKQHKSLPNSLG 780

DB 721 GDTIKGNLNSKALLSFLMSQHAKIKNQAITBTDNSVAVPOEKQHKSLPNSLG 780

QY 721 GDTIKGNLNSKALLSFLMSQHAKIKNQAITBTDNSVAVPOEKQHKSLPNSLG 780

DB 721 GDTIKGNLNSKALLSFLMSQHAKIKNQAITBTDNSVAVPOEKQHKSLPNSLG 780

QY 781 VSERLQRTTPAVSVKUNGHDQCONPDLDETTARFRVETHTQKTIQGNTKEKPSLIV 840

DB 781 VSERLQRTTPAVSVKUNGHDQCONPDLDETTARFRVETHTQKTIQGNTKEKPSLIV 840

QY 841 PLESQMTKEKKTGKEKENSREMAENHIGTEVILGKLUQHYTDSYLGFLPWECKYF 900

DB 841 PLESQMTKEKKTGKEKENSREMAENHIGTEVILGKLUQHYTDSYLGFLPWECKYF 900

QY 901 QDLDDEBSLSKIQLAYFFDSKOTGRQLK 928

DB 901 QDLDDEBSLSKIQLAYFFDSKOTGRQLK 928

QY RESULT 2

US-09-986-552-1

Sequence 1, Application US/0986552

Patent No. US2002010981A1

GENERAL INFORMATION:

APPLICANT: CANFIELD, William

TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSMAL HYDROLASES

FILE REFERENCE: 215089US77D17

CURRENT APPLICATION NUMBER: US/09/986, 552

CURRENT FILING DATE: 2001-11-09

PRIOR APPLICATION NUMBER: 09/635, 872

PRIOR FILING DATE: 2000-08-10

PRIOR APPLICATION NUMBER: 60/153, 831

PRIOR FILING DATE: 1999-09-14

NUMBER OF SEQ ID NOS: 52

SOFTWARE: PatentIn version 3.1

SEQ ID NO 1

LENGTH: 928

TYPE: PRT

ORGANISM: Homo sapiens

US-09-986-552-1

Query Match Best Local Similarity 100.0%; Score 4907; DB 9; Length 928; Matches 928; Conservative 0; Mismatches 0; Index 0; Gaps 0;

QY 1 MLFKLQLQRQTYCILCPKIDQYVFLGKQVNTVSAGFGEVULEWSRDQYHFLDSRDT 60

DB 1 MLFKLQLQRQTYCILCPKIDQYVFLGKQVNTVSAGFGEVULEWSRDQYHFLDSRDT 60

QY 61 AGKSFFNRCLCPMPDVYVTTWNGTDLBLKELQQREQMEEBQAMREIGNTTPTK 120

DB 61 AGKSFFNRCLCPMPDVYVTTWNGTDLBLKELQQREQMEEBQAMREIGNTTPTK 120

QY 121 KSEKOLECLLTHIKVPMVLDPALPANTIKDVKVPLSPFHASICFNVKPKNSTNV 180

DB 121 KSEKOLECLLTHIKVPMVLDPALPANTIKDVKVPLSPFHASICFNVKPKNSTNV 180

QY 181 SVVVFDSKVEDAHSGLKGNSRQTWRYGLTIDKEVPGVLMODIAFLSGFRPPFKET 240

DB 181 SVVVFDSKVEDAHSGLKGNSRQTWRYGLTIDKEVPGVLMODIAFLSGFRPPFKET 240

QY 181 SVVVFDSKVEDAHSGLKGNSRQTWRYGLTIDKEVPGVLMODIAFLSGFRPPFKET 240

DB 181 SVVVFDSKVEDAHSGLKGNSRQTWRYGLTIDKEVPGVLMODIAFLSGFRPPFKET 240

QY 241 NOLKTKLBNLSSKVKULQYSEASVALKLNPQFOELNKOTKONTIDGKELTSPA 300

DB 241 NOLKTKLBNLSSKVKULQYSEASVALKLNPQFOELNKOTKONTIDGKELTSPA 300

QY 301 YLWDLISLTSQSQDEDIISASFEDNEELRSYLSRSERHAPVNIFIVINGQISWLNL 360

DB 301 YLWDLISLTSQSQDEDIISASFEDNEELRSYLSRSERHAPVNIFIVINGQISWLNL 360

QY 361 DNRPVTITHQDFRNLSHLPSSPALESHIRIGLQSQFYIYLDDVMFGKDWPDD 420

DB 361 DNRPVTITHQDFRNLSHLPSSPALESHIRIGLQSQFYIYLDDVMFGKDWPDD 420

QY 421 YSHSKGQKVLTWVPVNAEGPGSWIKDGYDKACNSACWDGGDCSGNSGGSRYTAG 480

DB 421 YSHSKGQKVLTWVPVNAEGPGSWIKDGYDKACNSACWDGGDCSGNSGGSRYTAG 480

QY 481 GGGTGSISGVGHPMQFGGGINSVSYCNGCANSLWAKFCDCQACNVLSGFGDQDHF 540

DB 481 GGGTGSISGVGHPMQFGGGINSVSYCNGCANSLWAKFCDCQACNVLSGFGDQDHF 540

QY 541 HEYKVILLPNQHYYIPKGECIPLYSFAEVAKRGVEGAYDNPIRHASTANKWTHL 600

DB 541 HEYKVILLPNQHYYIPKGECIPLYSFAEVAKRGVEGAYDNPIRHASTANKWTHL 600

QY 601 IMMSGMATTIHLNTFONTDEEKFQKITVEVDTRGPUNKNSTAQKGYENLVSPTLLP 660

DB 601 IMMSGMATTIHLNTFONTDEEKFQKITVEVDTRGPUNKNSTAQKGYENLVSPTLLP 660

QY 651 EARTLFEPKKEGRPKERHVNSTRRAQEVKIPLNISLPLQALSLNTDQLEH 720

DB 651 EARTLFEPKKEGRPKERHVNSTRRAQEVKIPLNISLPLQALSLNTDQLEH 720

QY 721 GDTIKGNLNSKALLSFLMSQHAKIKNQAITBTDNSVAVPOEKQHKSLPNSLG 780

DB 721 GDTIKGNLNSKALLSFLMSQHAKIKNQAITBTDNSVAVPOEKQHKSLPNSLG 780

QY 781 VSERLQRTTPAVSVKUNGHDQCONPDLDETTARFRVETHTQKTIQGNTKEKPSLIV 840

DB 781 VSERLQRTTPAVSVKUNGHDQCONPDLDETTARFRVETHTQKTIQGNTKEKPSLIV 840

QY 841 PLESQMTKEKKTGKEKENSREMAENHIGTEVILGKLUQHYTDSYLGFLPWECKYF 900

DB 841 PLESQMTKEKKTGKEKENSREMAENHIGTEVILGKLUQHYTDSYLGFLPWECKYF 900

QY 901 QDLDDEBSLSKIQLAYFFDSKOTGRQLK 928

DB 901 QDLDDEBSLSKIQLAYFFDSKOTGRQLK 928

QY RESULT 3

US-10-023-888-4

; Sequence 4, Application US/10023888  
; Publication No. US20030119088A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: SOLUBLE GLCNAC PHOSPHOTRANSFERASE  
; FILE REFERENCE: 201515US77  
; CURRENT APPLICATION NUMBER: US/10/023,888  
; CURRENT FILING DATE: 2001-12-21  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 4  
; LENGTH: 928  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-10-023-888-4

Query Match 100.0%; Score 4907; DB 14; Length 928;  
Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;  
Matches 928; Conservative 0;

Qy 1 MLEFKLQLQRTYCLPMLVDPALPANTIKVPMVLQDFKQVPLGTVTTIVSAFQGEVWLEWSQDQHVLFDSYRDNI 60  
Db 1 MLFKLQLQRTYCLPMLVDPALPANTIKVPMVLQDFKQVPLGTVTTIVSAFQGEVWLEWSQDQHVLFDSYRDNI 60  
Qy 61 AGKSFQRNLCLPMLVDPVVTWNGTDLIELKEILQOVRQEMEERQKAMREELIGKNTTEPTK 120  
Db 61 AGKSFQRNLCLPMLVDPVVTWNGTDLIELKEILQOVRQEMEERQKAMREELIGKNTTEPTK 120  
Qy 121 KSEKQLECLLTCIKVPMVLQDFKQVPLGTVTTIVSAFQGEVWLEWSQDQHVLFDSYRDNI 180  
Db 121 KSEKQLECLLTCIKVPMVLQDFKQVPLGTVTTIVSAFQGEVWLEWSQDQHVLFDSYRDNI 180  
Qy 181 SVVVFDSSTDVKVEDAHSGLKGNSQTVWRGTYLTDEKVPGVGLVMDAFLSGFPPTKE 240  
Db 181 SVVVFDSSTDVKVEDAHSGLKGNSQTVWRGTYLTDEKVPGVGLVMDAFLSGFPPTKE 240  
Qy 241 NOLKTKLPENLSSKVKKLQLYSEASVALKLNPKDQELNKQTKOMTIDGKELTISPA 300  
Db 241 NOLKTKLPENLSSKVKKLQLYSEASVALKLNPKDQELNKQTKOMTIDGKELTISPA 300  
Qy 301 YJWLDLSAISQSQDEDISASRSPEDNEELRYSLSRTHREGLSQTIFYLNDVAFGKDWPDDF 360  
Db 301 YJWLDLSAISQSQDEDISASRSPEDNEELRYSLSRTHREGLSQTIFYLNDVAFGKDWPDDF 360  
Qy 361 DNPRVTITVHDYFRNLSHLPTFPSSPAIESHTHRLEGTSQTFIYLNDDVAFGKDWPDDF 420  
Db 361 DNPRVTITVHDYFRNLSHLPTFPSSPAIESHTHRLEGTSQTFIYLNDDVAFGKDWPDDF 420  
Qy 421 YHSKGQKQVLYLWVPPNCAEGCPGSWIKGDCKACNNASACWDGGDCSNSGSRSYIAG 480  
Db 421 YHSKGQKQVLYLWVPPNCAEGCPGSWIKGDCKACNNASACWDGGDCSNSGSRSYIAG 480  
Qy 481 GGTTGSIGVGHMQFGGGINSYCNGCANSWLAQKFCDAQCNVLQDFKQVPLGTVTTIVSAFQGEVWLEWSQDQHVLFDSYRDNI 540  
Db 481 GGTTGSIGVGHMQFGGGINSYCNGCANSWLAQKFCDAQCNVLQDFKQVPLGTVTTIVSAFQGEVWLEWSQDQHVLFDSYRDNI 540  
Qy 541 HEKYKVLIPNQTHYIIPKGECLPYSFAEVAKRGTEGAGSDNPITRHASIANKKTHI 600  
Db 541 HEKYKVLIPNQTHYIIPKGECLPYSFAEVAKRGTEGAGSDNPITRHASIANKKTHI 600  
Qy 601 IMHSGMATTIFNLFTONTDEEFKQITTEVDREGPKUNSTAOKGYENLVSPTILP 660  
Db 601 IMHSGMATTIFNLFTONTDEEFKQITTEVDREGPKUNSTAOKGYENLVSPTILP 660  
Qy 661 EABILFEDIPKSKRPPKRVNSTRRAQEVKVLPLNLSLPLQDQSLQNTDQIHL 720  
Db 661 EABILFEDIPKSKRPPKRVNSTRRAQEVKVLPLNLSLPLQDQSLQNTDQIHL 720  
Qy 721 GDITLKGNLNSKALRSPLMSOHAKKNOAIIITBTNSLIVAPRQEKQHKSILPSLG 780  
Db 721 GDITLKGNLNSKALRSPLMSOHAKKNOAIIITBTNSLIVAPRQEKQHKSILPSLG 780  
Qy 781 VSRQLQUTTPAVSVKUNGHDQGONPPLLETARFVETHQTKTIGNTKEKPPSLIV 840

RESULT 4  
; Sequence 4, Application US/10023889  
; Publication No. US20030124652A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: METHODS OF PRODUCING HIGH MANNOSE GLYCOPROTEINS IN COMPLEX CARBC  
; FILE REFERENCE: 203121US77  
; CURRENT APPLICATION NUMBER: US/10/023, 889  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 4  
; LENGTH: 928  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-10-023-889-4

Query Match 100.0%; Score 4907; DB 14; Length 928;  
Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;  
Matches 928; Conservative 0;

Qy 1 MLEFKLQLQRTYCLPMLVDPALPANTIKVPMVLQDFKQVPLGTVTTIVSAFQGEVWLEWSQDQHVLFDSYRDNI 60  
Db 1 MLFKLQLQRTYCLPMLVDPALPANTIKVPMVLQDFKQVPLGTVTTIVSAFQGEVWLEWSQDQHVLFDSYRDNI 60  
Qy 61 AGKSFQRNLCLPMLVDPVVTWNGTDLIELKEILQOVRQEMEERQKAMREELIGKNTTEPTK 120  
Db 61 AGKSFQRNLCLPMLVDPVVTWNGTDLIELKEILQOVRQEMEERQKAMREELIGKNTTEPTK 120  
Qy 121 KSEKQLECLLTCIKVPMVLQDFKQVPLGTVTTIVSAFQGEVWLEWSQDQHVLFDSYRDNI 180  
Db 121 KSEKQLECLLTCIKVPMVLQDFKQVPLGTVTTIVSAFQGEVWLEWSQDQHVLFDSYRDNI 180  
Qy 181 SVVVFDSSTDVKVEDAHSGLKGNSQTVWRGTYLTDEKVPGVGLVMDAFLSGFPPTKE 240  
Db 181 SVVVFDSSTDVKVEDAHSGLKGNSQTVWRGTYLTDEKVPGVGLVMDAFLSGFPPTKE 240  
Qy 241 NOLKTKLPENLSSKVKKLQLYSEASVALKLNPKDQELNKQTKOMTIDGKELTISPA 300  
Db 241 NOLKTKLPENLSSKVKKLQLYSEASVALKLNPKDQELNKQTKOMTIDGKELTISPA 300  
Qy 301 YJWLDLSAISQSQDEDISASRSPEDNEELRYSLSRTHREGLSQTIFYLNDVAFGKDWPDDF 360  
Db 301 YJWLDLSAISQSQDEDISASRSPEDNEELRYSLSRTHREGLSQTIFYLNDVAFGKDWPDDF 360  
Qy 361 DNPRVTITVHDYFRNLSHLPTFPSSPAIESHTHRLEGTSQTFIYLNDDVAFGKDWPDDF 420  
Db 361 DNPRVTITVHDYFRNLSHLPTFPSSPAIESHTHRLEGTSQTFIYLNDDVAFGKDWPDDF 420  
Qy 421 YHSKGQKQVLYLWVPPNCAEGCPGSWIKGDCKACNNASACWDGGDCSNSGSRSYIAG 480  
Db 421 YHSKGQKQVLYLWVPPNCAEGCPGSWIKGDCKACNNASACWDGGDCSNSGSRSYIAG 480  
Qy 481 GGTTGSIGVGHMQFGGGINSYCNGCANSWLAQKFCDAQCNVLQDFKQVPLGTVTTIVSAFQGEVWLEWSQDQHVLFDSYRDNI 540  
Db 481 GGTTGSIGVGHMQFGGGINSYCNGCANSWLAQKFCDAQCNVLQDFKQVPLGTVTTIVSAFQGEVWLEWSQDQHVLFDSYRDNI 540  
Qy 541 HEKYKVLIPNQTHYIIPKGECLPYSFAEVAKRGTEGAGSDNPITRHASIANKKTHI 600



QY 121 KSEKOLECILTHCKVPMVLDPALPANTIKDPSLPSHSASDFNVAKPKNSTNV 180 ; TYPE: PRT  
 Db 121 KSEKOLECILTHCKVPMVLDPALPANTIKDPSLPSHSASDFNVAKPKNSTNV 180 ; ORGANISM: Homo sapiens  
 QY 181 SVVVFDSITKDVEDAHSGLKGNSRQTWGRGILTDXKVPGLVIMODAFLAFSGFPPTK 240 ; US-10-023-894-4  
 Db 181 SVVVFDSITKDVEDAHSGLKGNSRQTWGRGILTDXKVPGLVIMODAFLAFSGFPPTK 240 ;  
 QY 241 NQKTKUPENLSKVVKVQLQLYEASVALKLNNPKDQELNKOTKONTIDKELTISPA 300 ;  
 Db 241 NQKTKUPENLSKVVKVQLQLYEASVALKLNNPKDQELNKOTKONTIDKELTISPA 300 ;  
 QY 301 YLLWDLSAISOSKODEDISASRFEDNEBLLRYSLRSIRHAPWNNTFVINGOISWLNL 360 ;  
 Db 301 YLLWDLSAISOSKODEDISASRFEDNEBLLRYSLRSIRHAPWNNTFVINGOISWLNL 360 ;  
 QY 361 DNPRTVTHQDVERNLSPHTSSPALESHIRIEGUSQFIFYLNDWFGKDWPDDF 420 ;  
 Db 361 DNPRTVTHQDVERNLSPHTSSPALESHIRIEGUSQFIFYLNDWFGKDWPDDF 420 ;  
 QY 421 YSHSKGQKYLTPVPGCAEGPGSMWKDGYCDKACINSAACDWDGGCSNCGSRYTAG 480 ;  
 Db 421 YSHSKGQKYLTPVPGCAEGPGSMWKDGYCDKACINSAACDWDGGCSNCGSRYTAG 480 ;  
 QY 481 GGGTGSISGVHPMOPGGGINSVSYCNGCANSWLAKECDQACNVLISCGFDAGDCQDHF 540 ;  
 Db 481 GGGTGSISGVHPMOPGGGINSVSYCNGCANSWLAKECDQACNVLISCGFDAGDCQDHF 540 ;  
 QY 541 HELYKVILPNOTHYIPKGECLPYFPEAVKRGVEGAYSDNPTRHASTANKWTHL 600 ;  
 Db 541 HELYKVILPNOTHYIPKGECLPYFPEAVKRGVEGAYSDNPTRHASTANKWTHL 600 ;  
 QY 601 IMESGMNATIHNLTFTONDEFKMOTIVTVDTRGPKLNSTAKGYENLVSPTLP 660 ;  
 Db 601 IMESGMNATIHNLTFTONDEFKMOTIVTVDTRGPKLNSTAKGYENLVSPTLP 660 ;  
 QY 661 EASTLFEDIPKERKPFKRHDVNSTRAQEEVKIPLVNTSLPKDQASLNTDQLEH 720 ;  
 Db 661 EASTLFEDIPKERKPFKRHDVNSTRAQEEVKIPLVNTSLPKDQASLNTDQLEH 720 ;  
 QY 721 GDTIIGMNLKSALLSFLMSOHAKKNAQITDETDLSVAPRKQHKSLPNSLG 780 ;  
 Db 721 GDTIIGMNLKSALLSFLMSOHAKKNAQITDETDLSVAPRKQHKSLPNSLG 780 ;  
 QY 781 VSRQLQTFPAVKVNGHDQCONPDLDETTARVETHTQKTTGKNTKEPKSLIV 840 ;  
 Db 781 VSRQLQTFPAVKVNGHDQCONPDLDETTARVETHTQKTTGKNTKEPKSLIV 840 ;  
 QY 841 PLESQMTKEKKIGKEKNSRMEENAHIGTEVILGKLOHYTDSYLGFLPWERKKYF 900 ;  
 Db 841 PLESQMTKEKKIGKEKNSRMEENAHIGTEVILGKLOHYTDSYLGFLPWERKKYF 900 ;  
 QY 901 QDLDDEBSIKTOLAYFTDSKNGRQLK 928 ;  
 Db 901 QDLDDEBSIKTOLAYFTDSKNGRQLK 928 ;  
 ;  
 RESULT 7 ;  
 US-10-023-894-4 ;  
 ; Sequence 4, Application US /10023894 ;  
 ; Publication No. US20030143669A1 ;  
 ; GENERAL INFORMATION :  
 ; APPLICANT: CANFIELD, William ;  
 ; APPLICANT: KORNBLD, Stuart ;  
 ; TITLE OF INVENTION: EXPRESSION OF LYSOSMAL HYDROLASE IN CELLS EXPRESSING PRO-N- ;  
 ; TITLE OF INVENTION: ACETYLGALACTOSAMINE-1-PHOSPHODIESTER ALPHA-N-ACETYL GLUCOSIMANIDAS ;  
 ; FILE REFERENCE: 217391877 ;  
 ; CURRENT APPLICATION NUMBER: US /10/023, 894 ;  
 ; CURRENT FILING DATE: 2001-12-21 ;  
 ; NUMBER OF SEQ ID NOS: 22 ;  
 ; SOFTWARE: Patentin version 3.1 ;  
 ; SEQ ID NO 4 ;  
 ; LENGTH: 928

Query Match 100.0%; Score 4907; DB 14; Length 928;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Query 1 MLFKULLQRQTYTCIHSRYGLYVFLGVVVTIVSAFORGEVWLSRSDYHVLFDSDYINTI 60 ;  
 Db 1 MLFKULLQRQTYTCIHSRYGLYVFLGVVVTIVSAFORGEVWLSRSDYHVLFDSDYINTI 60 ;  
 Qy 61 AGKSFRQRCLPMPIDVVYTWNGTDLIELKEQVROMEEFQKABEILGNTTEPTK 120 ;  
 Db 61 AGKSFRQRCLPMPIDVVYTWNGTDLIELKEQVROMEEFQKABEILGNTTEPTK 120 ;  
 Qy 181 SVVVFDSITKDVEDAHSGLKGNSRQTWGRGILTDXKVPGLVIMODAFLAFSGFPPTK 240 ;  
 Db 181 SVVVFDSITKDVEDAHSGLKGNSRQTWGRGILTDXKVPGLVIMODAFLAFSGFPPTK 240 ;  
 Qy 241 NQKTKUPENLSKVVKVQLQLYEASVALKLNNPKDQELNKOTKONTIDKELTISPA 300 ;  
 Db 241 NQKTKUPENLSKVVKVQLQLYEASVALKLNNPKDQELNKOTKONTIDKELTISPA 300 ;  
 Qy 301 YLLWDLSAISOSKODEDISASRFEDNEBLLRYSLRSIRHAPWNNTFVINGOISWLNL 360 ;  
 Db 301 YLLWDLSAISOSKODEDISASRFEDNEBLLRYSLRSIRHAPWNNTFVINGOISWLNL 360 ;  
 Qy 361 DNPRTVTHQDVERNLSPHTSSPALESHIRIEGUSQFIFYLNDWFGKDWPDDF 420 ;  
 Db 361 DNPRTVTHQDVERNLSPHTSSPALESHIRIEGUSQFIFYLNDWFGKDWPDDF 420 ;  
 Qy 421 YSHSKGQKYLTPVPGCAEGPGSMWKDGYCDKACINSAACDWDGGCSNCGSRYTAG 480 ;  
 Db 421 YSHSKGQKYLTPVPGCAEGPGSMWKDGYCDKACINSAACDWDGGCSNCGSRYTAG 480 ;  
 Qy 481 GGGTGSISGVHPMOPGGGINSVSYCNGCANSWLAKECDQACNVLISCGFDAGDCQDHF 540 ;  
 Db 481 GGGTGSISGVHPMOPGGGINSVSYCNGCANSWLAKECDQACNVLISCGFDAGDCQDHF 540 ;  
 Qy 541 HELYKVILPNOTHYIPKGECLPYFPEAVKRGVEGAYSDNPTRHASTANKWTHL 600 ;  
 Db 541 HELYKVILPNOTHYIPKGECLPYFPEAVKRGVEGAYSDNPTRHASTANKWTHL 600 ;  
 Qy 601 IMESGMNATIHNLTFTONDEFKMOTIVTVDTRGPKLNSTAKGYENLVSPTLP 660 ;  
 Db 601 IMESGMNATIHNLTFTONDEFKMOTIVTVDTRGPKLNSTAKGYENLVSPTLP 660 ;  
 Qy 661 EASTLFEDIPKERKPFKRHDVNSTRAQEEVKIPLVNTSLPKDQASLNTDQLEH 720 ;  
 Db 661 EASTLFEDIPKERKPFKRHDVNSTRAQEEVKIPLVNTSLPKDQASLNTDQLEH 720 ;  
 Qy 721 GDTIIGMNLKSALLSFLMSOHAKKNAQITDETDLSVAPRKQHKSLPNSLG 780 ;  
 Db 721 GDTIIGMNLKSALLSFLMSOHAKKNAQITDETDLSVAPRKQHKSLPNSLG 780 ;  
 Qy 781 VSRQLQTFPAVKVNGHDQCONPDLDETTARVETHTQKTTGKNTKEPKSLIV 840 ;  
 Db 781 VSRQLQTFPAVKVNGHDQCONPDLDETTARVETHTQKTTGKNTKEPKSLIV 840 ;  
 Qy 841 PLESQMTKEKKIGKEKNSRMEENAHIGTEVILGKLOHYTDSYLGFLPWERKKYF 900 ;  
 Db 841 PLESQMTKEKKIGKEKNSRMEENAHIGTEVILGKLOHYTDSYLGFLPWERKKYF 900 ;  
 Qy 901 QDLDDEBSIKTOLAYFTDSKNGRQLK 928 ;  
 Db 901 QDLDDEBSIKTOLAYFTDSKNGRQLK 928 ;

RESULT 8  
US-10-306-686-1  
; Sequence 1, Application US/1030686  
; Publication No. US2003014846031  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM  
; TITLE OF INVENTION: PHOSPHODISTER ALPHA-GLCNACase OF THE LYSOSMAL TARGETING PATHWAY  
; FILE REFERENCE: 233379US77D7V  
; CURRENT APPLICATION NUMBER: US/10/306, 686  
; CURRENT FILING DATE: 2002-11-29  
; PRIOR APPLICATION NUMBER: 09/636, 596  
; PRIOR FILING DATE: 2000-08-30  
; PRIORITY DATE: 1999-08-14  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO: 1  
; LENGTH: 928  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-10-306-686-1

Query Match 100.0%; Score 4907; DB 14; Length 928;  
Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;  
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFLKLLQDQTYTCILSHRYGLYTCFLGLWVTTSAFQGEVLEWSRDQYHVLFDSYRDI 60  
1 MLFKLLQDQTYTCILSHRYGLYTCFLGLWVTTSAFQGEVLEWSRDQYHVLFDSYRDI 60

QY 61 AGESFQNLCLPMLPMDIVYTWTNGTDLLELKSLQREQMEEQKMRRELGNNTTEPK 120  
61 AGESFQNLCLPMLPMDIVYTWTNGTDLLELKSLQREQMEEQKMRRELGNNTTEPK 120

Db 121 KSEKOLELILTHCKVPMVLDPALPANTIKDVKPSHSASIDINVAKPSTNV 180  
121 KSEKOLELILTHCKVPMVLDPALPANTIKDVKPSHSASIDINVAKPSTNV 180

QY 181 SVVFDSTDVKEDAHSGLKGNSRQTVWRGYLTDXKEVPGLVIMODAFLSGFPPTPKT 240  
181 SVVFDSTDVKEDAHSGLKGNSRQTVWRGYLTDXKEVPGLVIMODAFLSGFPPTPKT 240

Db 241 NOLKTKLPLNLSKVKLQLYSEASVALLKLNPKDFOELNKQTCKMKTGKETISPA 300  
241 NOLKTKLPLNLSKVKLQLYSEASVALLKLNPKDFOELNKQTCKMKTGKETISPA 300

QY 45 SRDQYHVLFDSYRDNIAKGSFONRLCIPMPDVYTWTNGTDLLELKSLQREQMEEQ 104  
45 SRDQYHVLFDSYRDNIAKGSFONRLCIPMPDVYTWTNGTDLLELKSLQREQMEEQ 94

Db 35 SRDQYHVLFDSYRDNIAKGSFONRLCIPMPDVYTWTNGTDLLELKSLQREQMEEQ 94  
35 SRDQYHVLFDSYRDNIAKGSFONRLCIPMPDVYTWTNGTDLLELKSLQREQMEEQ 94

QY 105 KAMRELLGKONTBEPKAKSKEKOLECUTHCICKPMSVLDPALPANTIKDVKPSHS 164  
95 KAMRELLGKONTBEPKAKSKEKOLECUTHCICKPMSVLDPALPANTIKDVKPSHS 154

Db 165 SDTFNVAKEPNPSTNVSVVFDSTDVKEDAHSGLKGNSRQTVWRGYLTDXKEVPGLVIM 224  
165 SDTFNVAKEPNPSTNVSVVFDSTDVKEDAHSGLKGNSRQTVWRGYLTDXKEVPGLVIM 224

QY 361 DNPRVTITWHDYFRNTLHPTSSPALESHTHRIGLSQKPYIYNDVMFGKDWPDF 420  
361 DNPRVTITWHDYFRNTLHPTSSPALESHTHRIGLSQKPYIYNDVMFGKDWPDF 420

Db 301 YLWLDSLAIOSKQDEPDISASFEDNEELRSYRSIRRHPWNRFNIFTINGQFSPWNL 360  
301 YLWLDSLAIOSKQDEPDISASFEDNEELRSYRSIRRHPWNRFNIFTINGQFSPWNL 360

QY 421 YHSKQKQVYLITWPVPVNAEGCGSWIKGDYKCDKACINSACDGGCSGGSRVYAG 480  
421 YHSKQKQVYLITWPVPVNAEGCGSWIKGDYKCDKACINSACDGGCSGGSRVYAG 480

Db 421 YHSKQKQVYLITWPVPVNAEGCGSWIKGDYKCDKACINSACDGGCSGGSRVYAG 480  
421 YHSKQKQVYLITWPVPVNAEGCGSWIKGDYKCDKACINSACDGGCSGGSRVYAG 480

QY 481 GGGTSGSISGVHMFQGGGINSVSYNCNGCANSWLAQDFCDOACNVISCGFDAGDGQDH 540  
481 GGGTSGSISGVHMFQGGGINSVSYNCNGCANSWLAQDFCDOACNVISCGFDAGDGQDH 540

Db 431 GGCTGSGISGVHMFQGGGINSVSYNCNGCANSWLAQDFCDOACNVISCGFDAGDGQDH 540  
431 GGCTGSGISGVHMFQGGGINSVSYNCNGCANSWLAQDFCDOACNVISCGFDAGDGQDH 540

QY 541 HEIYKVILLPNOHYTPRGCECPYRFAEVAKRGYEGQAYEDNPIRHASANKMTHL 600  
541 HEIYKVILLPNOHYTPRGCECPYRFAEVAKRGYEGQAYEDNPIRHASANKMTHL 600

Db 601 IMMSGMNTATHENLTQINTDBEFKQHQTIVVTDTEBGPKLNSTACKGYENLVSPTILP 660  
601 IMMSGMNTATHENLTQINTDBEFKQHQTIVVTDTEBGPKLNSTACKGYENLVSPTILP 660

Db 601 IMMSGMNTATHENLTQINTDBEFKQHQTIVVTDTEBGPKLNSTACKGYENLVSPTILP 660  
601 IMMSGMNTATHENLTQINTDBEFKQHQTIVVTDTEBGPKLNSTACKGYENLVSPTILP 660

QY 651 EAELFELDIPKECRPKPKRKHVNSTRQAQEVKIPVNLSLPKQALQSLNTDQLEH 720  
651 EAELFELDIPKECRPKPKRKHVNSTRQAQEVKIPVNLSLPKQALQSLNTDQLEH 720

Db 651 EAELFELDIPKECRPKPKRKHVNSTRQAQEVKIPVNLSLPKQALQSLNTDQLEH 720  
651 EAELFELDIPKECRPKPKRKHVNSTRQAQEVKIPVNLSLPKQALQSLNTDQLEH 720

QY 721 GDTIKGYNLKSALLRFLMSQHAKIKNOAIIITDNDLVAPEKQVKHSILPNSIG 780  
721 GDTIKGYNLKSALLRFLMSQHAKIKNOAIIITDNDLVAPEKQVKHSILPNSIG 780

Db 781 VSERLQRUTFPASVVKVNGHDOCQNPDLTETARRFVETHQKTTGGNVTEKEPPSLIV 840  
781 VSERLQRUTFPASVVKVNGHDOCQNPDLTETARRFVETHQKTTGGNVTEKEPPSLIV 840

QY 841 PLEQMTKEKKINGKEKENSRMVEENAENHIGTYEVILGRKQHQTYSYLGFLPWEKKYF 900  
841 PLEQMTKEKKINGKEKENSRMVEENAENHIGTYEVILGRKQHQTYSYLGFLPWEKKYF 900

Db 901 QDLDDEESLSKTOLAYFDSKQGRQLK 928  
901 QDLDDEESLSKTOLAYFDSKQGRQLK 928

Db 901 QDLDDEESLSKTOLAYFDSKQGRQLK 928  
901 QDLDDEESLSKTOLAYFDSKQGRQLK 928

RESULT 9  
US-10-023-888-2  
; Sequence 2, Application US/10023888  
; Publication No. US2003019088A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: SOLUBLE GLCNAC PHOSPHOTRANSFERASE  
; FILE REFERENCE: 200515US77  
; CURRENT APPLICATION NUMBER: US/10/023, 888  
; CURRENT FILING DATE: 2001-12-21  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO: 2  
; LENGTH: 1199  
; TYPE: PRT  
; ORGANISM: hybrid  
; US-10-023-888-2

Query Match 94.4%; Score 4630; DB 14; Length 1199;  
Best Local Similarity 99.3%; Pred. No. 0; Mismatches 2; Indels 4; Gaps 0; Gaps 0;  
Matches 875; Conservative 2; Mismatches 4; Indels 4; Gaps 0; Gaps 0;

QY 45 SRDQYHVLFDSYRDNIAKGSFONRLCIPMPDVYTWTNGTDLLELKSLQREQMEEQ 104  
45 SRDQYHVLFDSYRDNIAKGSFONRLCIPMPDVYTWTNGTDLLELKSLQREQMEEQ 94

Db 35 SRDQYHVLFDSYRDNIAKGSFONRLCIPMPDVYTWTNGTDLLELKSLQREQMEEQ 94  
35 SRDQYHVLFDSYRDNIAKGSFONRLCIPMPDVYTWTNGTDLLELKSLQREQMEEQ 94

QY 105 KAMRELLGKONTBEPKAKSKEKOLECUTHCICKPMSVLDPALPANTIKDVKPSHS 164  
95 KAMRELLGKONTBEPKAKSKEKOLECUTHCICKPMSVLDPALPANTIKDVKPSHS 154

Db 165 SDTFNVAKEPNPSTNVSVVFDSTDVKEDAHSGLKGNSRQTVWRGYLTDXKEVPGLVIM 224  
165 SDTFNVAKEPNPSTNVSVVFDSTDVKEDAHSGLKGNSRQTVWRGYLTDXKEVPGLVIM 224

QY 155 SDTFNVAKEPNPSTNVSVVFDSTDVKEDAHSGLKGNSRQTVWRGYLTDXKEVPGLVIM 214  
225 QDIAFLSGRPPTKETNOLKTKLPENTSSKVKLQLYSEASVALLKLNPKDFOELNKQT 284  
225 QDIAFLSGRPPTKETNOLKTKLPENTSSKVKLQLYSEASVALLKLNPKDFOELNKQT 284

Db 215 QDIAFLSGRPPTKETNOLKTKLPENTSSKVKLQLYSEASVALLKLNPKDFOELNKQT 274  
215 QDIAFLSGRPPTKETNOLKTKLPENTSSKVKLQLYSEASVALLKLNPKDFOELNKQT 274

Db 285 KKNMTIDKEKLTSPPAYLWDISAIOSKQDEPDISASFEDNEELRSYRSIRRHPWVR 344  
275 KKNMTIDKEKLTSPPAYLWDISAIOSKQDEPDISASFEDNEELRSYRSIRRHPWVR 334

Db 345 NIIFTNGQIPSMLNDPMPRTVTHDVFMLSHPTFSPATESHVRIEGLSKFY 404  
345 NIIFTNGQIPSMLNDPMPRTVTHDVFMLSHPTFSPATESHVRIEGLSKFY 404

Db 335 NIIFTNGQIPSMLNDPMPRTVTHDVFMLSHPTFSPATESHVRIEGLSKFY 394  
335 NIIFTNGQIPSMLNDPMPRTVTHDVFMLSHPTFSPATESHVRIEGLSKFY 394

Db 405 LNDWMFGKDVWDDPFSSHKSKQKVVITWPPVNCAGCGPGSWIKDQYCDKACNSACDW 464  
395 LNDWMFGKDVWDDPFSSHKSKQKVVITWPPVNCAGCGPGSWIKDQYCDKACNSACDW 454

Db 465 GGDGSGNSGSRVIAGGGATGGGSIAGGGATGGGSIAGGGATGGGINSVSYCQGCAWSLADKFCOACN 524  
455 GGDGSGNSGSRVIAGGGATGGGSIAGGGATGGGSIAGGGATGGGINSVSYCQGCAWSLADKFCOACN 514

Db 525 VLSCGGFDAQCGDQDHFLYKVILLPNOTHYIIPKBCCLPVFSFABEVAKRGVEGAVSDNP 584  
525 VLSCGGFDAQCGDQDHFLYKVILLPNOTHYIIPKBCCLPVFSFABEVAKRGVEGAVSDNP 584

Db 515 VLSCGFPRAGCQDHFLHYKVLPLNQTHYIPKGCLPYSPAEVAKRGVEGASDNP 574  
 Qy 585 IIRHASIANKWTKTHLIMSGNATIHLNPFONTDEBPKRQITVETDREGKINST 644  
 Db 575 IIRHASIANKWTKTHLIMSGNATIHLNPFONTDEBPKRQITVETDREGKINST 634  
 Qy 645 AOKGYENIUSPITLPEAEIPDIPKEPKRPFKRDVNSTRRAOBZEVKPLVNLSLP 704  
 Db 635 AOKGYENIUSPITLPEAEIPDIPKEPKRPFKRDVNSTRRAOBZEVKPLVNLSLP 694  
 Qy 705 KQDQLSINTLDIOLERHDITLGCKYNSKALLRSFLMSOHAKIKNQAITDENTSLVA 764  
 Db 695 KQDQLSINTLDIOLERHDITLGCKYNSKALLRSFLMSOHAKIKNQAITDENTSLVA 754  
 Qy 765 POKQVKISILNSLGYSERQRLTFAVSUKNNGDQGQNPDLDETTARFVETHQK 824  
 Db 755 POKQVKISILNSLGYSERQRLTFAVSUKNNGDQGQNPDLDETTARFVETHQK 814  
 Qy 825 TIGGNVTKERPPSLIVPLESOMTKEKLTGKEKENSMEENAHIGVTLGKUHQH 884  
 Db 815 TIGGNVTKERPPSLIVPLESOMTKEKLTGKEKENSMEENAHIGVTLGKUHQH 874  
 Qy 885 TDSYLGLPWEKKYFQDLDDEBESLKTOLAYFTDSRNAR 925  
 Db 875 TDSYLGLPWEKKYFQDLDDEBESLKTOLAYFTDSRNAR 915

RESULT 10  
 US-1-023-889-2  
 ; Sequence 2, Application US/10023889  
 ; Publication No. US20030124653A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CANFIELD, William  
 ; TITLE OF INVENTION: METHODS OF PRODUCING HIGH MANNOSE GLYCOPROTEINS IN COMPLEX CARBOHYDRATE RECEPTORS  
 ; FILE REFERENCE: 201510577  
 ; CURRENT APPLICATION NUMBER: US/10/023, 889  
 ; CURRENT FILING DATE: 2001-12-21  
 ; NUMBER OF SEQ ID NOS: 21

; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 2  
 ; LENGTH: 1199  
 ; TYPE: PRT  
 ; ORGANISM: hybrid

US-10-023-889-2

Query Match 94.4%; Score 4630; DB 14; Length 1199;  
 Best Local Similarity 99.3%; Pred. No. 0;  
 Matches 875; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 Qy 45 SRDQHVFDFSTRDNIAGKSFRNLCPMPIDVWYTWTNGDLELIKELQVREOMEEQ 104  
 Db 35 SRDQHVFDFSTRDNIAGKSFRNLCPMPIDVWYTWTNGDLELIKELQVREOMEEQ 94  
 Qy 105 KAMEBLIGKNTBPTKSKQBLCLTHICKIKPMLVDPALPANTIKDPSLYSFHSA 164  
 Db 95 KAMEBLIGKNTBPTKSKQBLCLTHICKIKPMLVDPALPANTIKDPSLYSFHSA 154  
 Qy 165 SDIFNVAKPKNPSINVVVFSTKVEDAHSGLLKNRSQTWRGILTDEKEVGLVLM 224  
 Db 155 SDIFNVAKPKNPSINVVVFSTKVEDAHSGLLKNRSQTWRGILTDEKEVGLVLM 214  
 Qy 225 QDIAFLSDFPPTKETKQBLCLTHICKIKPMLVDPALPANTIKDPSLYSFHSA 164  
 Db 215 QDIAFLSDFPPTKETKQBLCLTHICKIKPMLVDPALPANTIKDPSLYSFHSA 154  
 Qy 285 KQMTDICKELTISPAYLWDLSAISOKSQDEDISASRFNEEYLRSISIERHPWR 344  
 Db 275 KQMTDICKELTISPAYLWDLSAISOKSQDEDISASRFNEEYLRSISIERHPWR 334  
 Qy 345 NTIFVYNGQIPSMLNDPRTVTHDYFVRLSHLTFSSPAIESHHRIGLQSQFTV 404

Db 335 NIFLVNGQIPSMLNDPRTVTHDYFVRLSHLTFSSPAIESHHRIGLQSQFTV 394  
 Qy 405 LDNDVMFGDWDPPDYPSHSGCKVYLTWPVPCACPGSPSIKQYCDKACNNSDWD 464  
 Db 395 LDNDVMFGDWDPPDYPSHSGCKVYLTWPVPCACPGSPSIKQYCDKACNNSDWD 454  
 Qy 465 GGDGSQNSGSRVIAQGGGTGSTGVGHFWQFGGINNSVYCGCANSWLAJZKFCDAQN 524  
 Db 455 GGCGSGNSGGSRYAGGGTGSTGVGHFWQFGGINNSVYCGCANSWLAJZKFCDAQN 514  
 Qy 525 VLSCGFPRAGCQDHFLHYKVLPLNQTHYIPKGCLPYSPAEVAKRGVEGASDNP 584  
 Db 515 VLSCGFPRAGCQDHFLHYKVLPLNQTHYIPKGCLPYSPAEVAKRGVEGASDNP 574  
 Qy 585 IIRHASIANKWTKTHLIMSGNATIHLNPFONTDEBPKRQITVETDREGKINST 644  
 Db 575 IIRHASIANKWTKTHLIMSGNATIHLNPFONTDEBPKRQITVETDREGKINST 634  
 Qy 645 AOKGYENIUSPITLPEAEIPDIPKEPKRPFKRDVNSTRRAOBZEVKPLVNLSLP 704  
 Db 635 AOKGYENIUSPITLPEAEIPDIPKEPKRPFKRDVNSTRRAOBZEVKPLVNLSLP 694  
 Qy 705 KQDQLSINTLDIOLERHDITLGCKYNSKALLRSFLMSOHAKIKNQAITDENTSLVA 764  
 Db 695 KQDQLSINTLDIOLERHDITLGCKYNSKALLRSFLMSOHAKIKNQAITDENTSLVA 754  
 Qy 765 POKQVKISILNSLGYSERQRLTFAVSUKNNGDQGQNPDLDETTARFVETHQK 824  
 Db 755 POKQVKISILNSLGYSERQRLTFAVSUKNNGDQGQNPDLDETTARFVETHQK 814  
 Qy 825 TIGGNVTKERPPSLIVPLESOMTKEKLTGKEKENSMEENAHIGVTLGKUHQH 884  
 Db 815 TIGGNVTKERPPSLIVPLESOMTKEKLTGKEKENSMEENAHIGVTLGKUHQH 874  
 Qy 885 TDSYLGLPWEKKYFQDLDDEBESLKTOLAYFTDSRNAR 925  
 Db 875 TDSYLGLPWEKKYFQDLDDEBESLKTOLAYFTDSRNAR 915

RESULT 11  
 US-10-023-890-2  
 ; Sequence 2, Application US/10023890  
 ; Publication No. US20030124653A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CANFIELD, William  
 ; TITLE OF INVENTION: METHOD OF PRODUCING GLYCOPROTEINS HAVING REDUCED COMPLEX CARBOHYDRATE RECEPTORS  
 ; FILE REFERENCE: 201510577  
 ; CURRENT APPLICATION NUMBER: US/10/023, 890  
 ; CURRENT FILING DATE: 2001-12-21  
 ; NUMBER OF SEQ ID NOS: 21

; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 2  
 ; LENGTH: 1199  
 ; TYPE: PRT  
 ; ORGANISM: hybrid

US-10-023-890-2

Query Match 94.4%; Score 4630; DB 14; Length 1199;  
 Best Local Similarity 99.3%; Pred. No. 0;  
 Matches 875; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 Qy 45 SRDQHVFDFSTRDNIAGKSFRNLCPMPIDVWYTWTNGDLELIKELQVREOMEEQ 104  
 Db 35 SRDQHVFDFSTRDNIAGKSFRNLCPMPIDVWYTWTNGDLELIKELQVREOMEEQ 94  
 Qy 105 KAMEBLIGKNTBPTKSKQBLCLTHICKIKPMLVDPALPANTIKDPSLYSFHSA 164  
 Db 95 KAMEBLIGKNTBPTKSKQBLCLTHICKIKPMLVDPALPANTIKDPSLYSFHSA 154  
 Qy 165 SDIFNVAKPKNPSINVVVFSTKVEDAHSGLLKNRSQTWRGILTDEKEVGLVLM 224  
 Db 155 SDIFNVAKPKNPSINVVVFSTKVEDAHSGLLKNRSQTWRGILTDEKEVGLVLM 214

QY 225 QDIAFLSGRPPTPEKTKOIKLKPENLSSKVULQLYSEASVALLKUNPKDFOEINKT 284  
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 215 QDIAFLSGRPPTPEKTKOIKLKPENLSSKVULQLYSEASVALLKUNPKDFOEINKT 274  
 QY 285 KRNMTIDGKELTSPAYLIWLDISAISSQKDEDISASRFEDNEELYSLSRSLISTERHAWVR 344  
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 275 KRNMTIDGKELTSPAYLIWLDISAISSQKDEDISASRFEDNEELYSLSRSLISTERHAWVR 334  
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 345 NIFTVTNGOIPSWNLNDNPRTVITVHQDFRNULSHLPTFSSPAIESHRIEGLSOKFY 404  
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 335 NIFTVTNGOIPSWNLNDNPRTVITVHQDFRNULSHLPTFSSPAIESHRIEGLSOKFY 394  
 QY 405 LNDVMFEGKDWPDPFYHSKSGKOVYLTPWVNCAGCPGSNIKDGYCDKACNNACWD 464  
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 395 LNDVMFEGKDWPDPFYHSKSGKOVYLTPWVNCAGCPGSNIKDGYCDKACNNACWD 454  
 QY 455 GGDGSGNSGGSRTAGGGTGTGIGVQBWQFGGINSVSYCNQGANSWLADKFCDQACN 514  
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 525 VLSQGFDAGDCGQDHFHLYKVILLPQTHYIPKGCLPLPSFAVAKRGVEGAVSDNP 584  
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 515 VLSQGFDAGDCGQDHFHLYKVILLPQTHYIPKGCLPLPSFAVAKRGVEGAVSDNP 574  
 QY 585 IIRHASIANKWKTHLIMSGMATTIHFNLTFONTNDEEPRQIIVETDREGPLNST 634  
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 645 AOKGYENLVSPTILPEAEILFEDIPIKEKRFPKFKFRDNDSTRRAEVKPLVNISLP 704  
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 635 AOKGYENLVSPTILPEAEILFEDIPIKEKRFPKFKFRDNDSTRRAEVKPLVNISLP 694  
 QY 755 POEKQVHSILPNSLGVYERLORLTFFAVSKVNGHDQGONPLDLETTARFRVHTQK 824  
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 755 POEKQVHSILPNSLGVYERLORLTFFAVSKVNGHDQGONPLDLETTARFRVHTQK 814  
 QY 825 TIGANTVKTEKKPSSLIVPLESOMTKEKITGKEKENSMEENAHVGTEVLGRLQHY 884  
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 815 TIGANTVKTEKKPSSLIVPLESOMTKEKITGKEKENSMEENAHVGTEVLGRLQHY 874  
 QY 885 TDSYLGFLPWEKKYFLDLDBEESEESKIQLAYFTDSKGR 925  
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 875 TDSYLGFLPWEKKYFLDLDBEESEESKIQLAYFTDSKGR 915  
 RESULT 12  
 US-10-024-197-2  
 ; Sequence 2, Application US/10024197  
 ; Publication No. US20030133924A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CANFIELD, William  
 ; TITLE OF INVENTION: HIGHLY PHOSPHORYLATED ACID BETA-GLUCOCEREBROSIDASE AND METHODS OF  
 ; FILE REFERENCE: 20794150  
 ; CURRENT APPLICATION NUMBER: US/10/024,197  
 ; CURRENT FILING DATE: 2001-12-21  
 ; NUMBER OF SEQ ID NOS: 27  
 ; SOFTWARE: Patentin version 3.1  
 ; SEQ ID NO 2  
 ; LENGTH: 1199  
 ; TYPE: PRT  
 ; ORGANISM: hybrid  
 ; US-10-024-197-2  
 Query Match 94.4%; Score 4630; DB 14; Length 1199;  
 Best Local Similarity 99.3%; Pred. No. 0; Matches 875; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 ; CURRENT FILING DATE: 2001-12-21  
 ;  
 QY 45 SRDOHVLFDSYRDNTAKSFONRLICLMPIDPVYTAWNGTIDELLAKELQOVRQEQQ 104  
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 35 SRDOHVLFDSYRDNTAKSFONRLICLMPIDPVYTAWNGTIDELLAKELQOVRQEQQ 94  
 QY 105 KAMERILGNTTEKTKSBRKOLCCLTHCKUPMLVDPALANTILKDVLSYRPHSA 164  
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 95 KAMERILGNTTEKTKSBRKOLCCLTHCKUPMLVDPALANTILKDVLSYRPHSA 154  
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 225 QDIAFLSGRPPTPEKTKOIKLKPENLSSKVULQLYSEASVALLKUNPKDFOEINKT 284  
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 215 QDIAFLSGRPPTPEKTKOIKLKPENLSSKVULQLYSEASVALLKUNPKDFOEINKT 274  
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 275 KRNMTIDGKELTSPAYLIWLDISAISSQKDEDISASRFEDNEELYSLSRSLISTERHAWVR 334  
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 345 NIFTVTNGOIPSWNLNDNPRTVITVHQDFRNULSHLPTFSSPAIESHRIEGLSOKFY 404  
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 335 NIFTVTNGOIPSWNLNDNPRTVITVHQDFRNULSHLPTFSSPAIESHRIEGLSOKFY 394  
 QY 405 LNDVMFEGKDWPDPFYHSKSGKOVYLTPWVNCAGCPGSNIKDGYCDKACNNACWD 464  
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 395 LNDVMFEGKDWPDPFYHSKSGKOVYLTPWVNCAGCPGSNIKDGYCDKACNNACWD 454  
 QY 455 GGDGSGNSGGSRTAGGGTGTGIGVQBWQFGGINSVSYCNQGANSWLADKFCDQACN 514  
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 525 VLSQGFDAGDCGQDHFHLYKVILLPQTHYIPKGCLPLPSFAVAKRGVEGAVSDNP 584  
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 515 VLSQGFDAGDCGQDHFHLYKVILLPQTHYIPKGCLPLPSFAVAKRGVEGAVSDNP 574  
 QY 585 IIRHASIANKWKTHLIMSGMATTIHFNLTFONTNDEEPRQIIVETDREGPLNST 634  
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 645 AOKGYENLVSPTILPEAEILFEDIPIKEKRFPKFKFRDNDSTRRAEVKPLVNISLP 704  
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 635 AOKGYENLVSPTILPEAEILFEDIPIKEKRFPKFKFRDNDSTRRAEVKPLVNISLP 694  
 QY 755 POEKQVHSILPNSLGVYERLORLTFFAVSKVNGHDQGONPLDLETTARFRVHTQK 824  
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 755 POEKQVHSILPNSLGVYERLORLTFFAVSKVNGHDQGONPLDLETTARFRVHTQK 814  
 QY 825 TIGANTVKTEKKPSSLIVPLESOMTKEKITGKEKENSMEENAHVGTEVLGRLQHY 884  
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 815 TIGANTVKTEKKPSSLIVPLESOMTKEKITGKEKENSMEENAHVGTEVLGRLQHY 874  
 QY 885 TDSYLGFLPWEKKYFLDLDBEESEESKIQLAYFTDSKGR 925  
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 875 TDSYLGFLPWEKKYFLDLDBEESEESKIQLAYFTDSKGR 915  
 RESULT 13  
 US-10-023-894-2  
 ; Sequence 2, Application US/10023894  
 ; Publication No. US20030143669A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CANFIELD, William  
 ; APPLICANT: KORNFIELD, Stuart  
 ; TITLE OF INVENTION: EXPRESSION OF LYOSOMAL HYDROLASE IN CELLS EXPRESSING PRO-N-  
 ; FILE REFERENCE: 217139US77  
 ; CURRENT APPLICATION NUMBER: US/10/023,894  
 ; CURRENT FILING DATE: 2001-12-21  
 ;

RESULT 14  
NUMBER OF SEQ ID NOS: 22 ;  
SOFTWARE: PatentIn version 3.1 ;  
SEQ ID NO: 2 ;  
LENGTH: 1199 ;  
TYPE: PRT ;  
ORGANISM: hybrid ;  
US-10-023-894-2 ;

Query Match 94.4%; Score 4630; DB 14; Length 1199;  
Best Local Similarity 99.3%; Pred. No. 0; Mismatches 4; Indels 0; Gaps 0;  
Matches 875; Conservative 2; MisMatches 4; Indels 0; Gaps 0;

QY 45 SRDQYHVLFDSDYDNAGSKSFKRLCLPMLPIDVYTWTNGTDLLELKLQYREQMEEQ 104  
Db 35 SRDQYHVLFDSDYDNAGSKSFKRLCLPMLPIDVYTWTNGTDLLELKLQYREQMEEQ 94

QY 105 KARDELLKGKNTTEPTKSKBOKLCLLTICIKUMLVDPALPANTIKDLSLPSHSA 164  
Db 95 KAMRELLKGKNTTEPTKSKBOKLCLLTICIKUMLVDPALPANTIKDLSLPSHSA 154

QY 165 SDIENVAQPKNPSTNVSUVWFSDSTKDVEDAHSGLKGSNSQTWRGMLTDKEVPGIVLM 224  
Db 155 SDIENVAQPKNPSTNVSUVWFSDSTKDVEDAHSGLKGSNSQTWRGMLTDKEVPGIVLM 214

QY 225 QDIAFLSLSPPTKETNOLKTLPENNSKVKULLQYSEASVALKKUNPKFOEINKOT 284  
Db 215 QDIAFLSLSPPTKETNOLKTLPENNSKVKULLQYSEASVALKKUNPKFOEINKOT 274

QY 285 KKNMTIDKELTISPAVILWDSIASPKDQEISASRFEEDNEELRSRSTERHAWR 344  
Db 275 KKNMTIDKELTISPAVILWDSIASPKDQEISASRFEEDNEELRSRSTERHAWR 334

QY 345 NIIFTVNGQIPSMWNLDMPRVTWTHQDFVRFNJSHLPPFSSPAESHTHRFGJSQKPIY 404  
Db 335 NIIFTVNGQIPSMWNLDMPRVTWTHQDFVRFNJSHLPPFSSPAESHTHRFGJSQKPIY 394

QY 405 LNDDVWMFGKDVWDDPFYSHSKQKVYLTPVNCAGECPGSWTKGYCDKACCNACDWD 464  
Db 395 LNDDVWMFGKDVWDDPFYSHSKQKVYLTPVNCAGECPGSWTKGYCDKACCNACDWD 454

QY 465 GGCQSGNSGSGSRYIAGGGGTGSGTGVGHMWFQGGINNSYCHQGANSWLAQKFCDACN 524  
Db 455 GGCQSGNSGSGSRYIAGGGGTGSGTGVGHMWFQGGINNSYCHQGANSWLAQKFCDACN 514

QY 525 VLSGCCFPDAGCGDQPHFLPHLYKVTLPLNQTHYIIPKGECLPYSFAYEVAKRGVEGAYGSDNP 584  
Db 515 VLSCGPFDADCGDQPHFLPHLYKVTLPLNQTHYIIPKGECLPYSFAYEVAKRGVEGAYGSDNP 574

QY 585 IIRHASICWKIHLLHSGMMATIHFNLONTDEBFKQITVEVDTRBGKJNST 644  
Db 575 IIRHASICWKIHLLHSGMMATIHFNLONTDEBFKQITVEVDTRBGKJNST 634

QY 645 AQQGYENLVSPITLPEAIIIFEDIPKERKPKERKHDVNSTRRAQEVKIPILVNSILP 704  
Db 635 AQQGYENLVSPITLPEAIIIFEDIPKERKPKERKHDVNSTRRAQEVKIPILVNSILP 694

QY 705 KDAQSLINTLQDQHEGDTLKGNLNSKALSRSFMSQHAKKNAITBETDSLVA 764  
Db 695 KDAQSLINTLQDQHEGDTLKGNLNSKALSRSFMSQHAKKNAITBETDSLVA 754

QY 765 PQEKQWHAISLPLNSLGYSERLQLTTPAVSKVKGHDQGQNPLDLETTARFRVETHQK 824  
Db 755 PQEKQWHAISLPLNSLGYSERLQLTTPAVSKVKGHDQGQNPLDLETTARFRVETHQK 814

QY 825 TIGGVNTKCKPSLIVPLESOMTKEKKITGKEKENSRSMEAENHGTEVLFGRKLOHY 884  
Db 815 TIGGVNTKCKPSLIVPLESOMTKEKKITGKEKENSRSMEAENHGTEVLFGRKLOHY 874

QY 885 TDSVYGFPLWEEKCKQFDLIDEERSKQOLAYFDSRNGR 925  
Db 875 TDSVYGFPLWEEKCKQFDLIDEERSKQOLAYFDSRNGR 915

Query Match 91.2%; Score 4473; DB 15; Length 845;  
Best Local Similarity 99.8%; Pred. No. 0; Mismatches 1; Indels 1;  
Matches 845; Conservative 1; MisMatches 1; Indels 1;

QY 1 MDFKLIQROTYTCVLSRHYGLYXCFGLGVVUTWSAOFGEVWLEWSRDOYI 61  
Db 1 MDFKLIQROTYTCVLSRHYGLYXCFGLGVVUTWSAOFGEVWLEWSRDOYI

QY 61 AKSFONRLCLPMLPIDVYTWTNGTDLLELKLQYREQMEEQKARRE 61  
Db 61 AKSFONRLCLPMLPIDVYTWTNGTDLLELKLQYREQMEEQKARRE

QY 121 KSFKQLECLLTCIKUMLVDPALPANTIKDPSLPSHSAISFNT 121  
Db 121 KSFKQLECLLTCIKUMLVDPALPANTIKDPSLPSHSAISFNT

QY 181 SVWVFSDSTKDVEDAHSGLKGSNSQTWRGMLTDKEVPGIVLWMDLAFT 181  
Db 181 SVWVFSDSTKDVEDAHSGLKGSNSQTWRGMLTDKEVPGIVLWMDLAFT

QY 241 NQKLTLPENNSKVKULLQYSEASVALKKUNPKFOEINKOTKQMT 241 NQKLTLPENNSKVKULLQYSEASVALKKUNPKFOEINKOTKQMT

QY 301 YIWLWDISAISPKDQEISASRFEEDNEELRSRSTERHAWRNFTIV 301 YIWLWDISAISPKDQEISASRFEEDNEELRSRSTERHAWRNFTIV

US-10-120-801-88 ;  
TYPE: PRT ;  
ORGANISM: human ;  
SEQ ID NO: 88 ;  
LENGTH: 847 ;  
; ORGANISM: human ;  
US-10-120-801-88 ;

RESULT 15  
Sequence 88, Application US/10120801 ;  
Publication No. US2003020383A1 ;  
GENERAL INFORMATION:  
APPLICANT: Guo, Xiaojia  
APPLICANT: Pena, Carol  
APPLICANT: Shimkets, Richard  
APPLICANT: Radigaru, Murliidhara  
APPLICANT: Kekuda, Ramesh  
APPLICANT: Spytek, Kimberly  
APPLICANT: Mehraban, Fuad  
APPLICANT: Topper, James N  
APPLICANT: Nalyankar, Uriel  
APPLICANT: Wasserman, Scott  
APPLICANT: Edinger, Shlomit  
APPLICANT: Smitsion, Gleinda  
APPLICANT: Gunther, Erik  
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 21402-340  
CURRENT APPLICATION NUMBER: US/10/120,801  
CURRENT FILING DATE: 2002-04-11  
PRIOR APPLICATION NUMBER: 60/285748  
PRIOR FILING DATE: 2001-04-23  
PRIOR APPLICATION NUMBER: 60/286068  
PRIOR FILING DATE: 2001-04-24  
PRIOR APPLICATION NUMBER: 60/286292  
PRIOR FILING DATE: 2001-04-25  
PRIOR APPLICATION NUMBER: 60/288334  
PRIOR FILING DATE: 2001-05-03  
PRIOR APPLICATION NUMBER: 60/291241  
PRIOR FILING DATE: 2001-05-16  
PRIOR APPLICATION NUMBER: 60/322284  
PRIOR FILING DATE: 2001-09-14  
PRIOR APPLICATION NUMBER: 60/285609  
PRIOR FILING DATE: 2001-04-20  
NUMBER OF SEQ ID NOS: 155  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO: 88  
LENGTH: 847  
; TYPE: PRT  
; ORGANISM: human  
US-10-120-801-88 ;

Query Match 91.2%; Score 4473; DB 15; Length 845;  
Best Local Similarity 99.8%; Pred. No. 0; Mismatches 1; Indels 1;  
Matches 845; Conservative 1; MisMatches 1; Indels 1;

QY 1 MDFKLIQROTYTCVLSRHYGLYXCFGLGVVUTWSAOFGEVWLEWSRDOYI 61  
Db 1 MDFKLIQROTYTCVLSRHYGLYXCFGLGVVUTWSAOFGEVWLEWSRDOYI

QY 61 AKSFONRLCLPMLPIDVYTWTNGTDLLELKLQYREQMEEQKARRE 61  
Db 61 AKSFONRLCLPMLPIDVYTWTNGTDLLELKLQYREQMEEQKARRE

QY 121 KSFKQLECLLTCIKUMLVDPALPANTIKDPSLPSHSAISFNT 121  
Db 121 KSFKQLECLLTCIKUMLVDPALPANTIKDPSLPSHSAISFNT

QY 181 SVWVFSDSTKDVEDAHSGLKGSNSQTWRGMLTDKEVPGIVLWMDLAFT 181 SVWVFSDSTKDVEDAHSGLKGSNSQTWRGMLTDKEVPGIVLWMDLAFT

QY 241 NQKLTLPENNSKVKULLQYSEASVALKKUNPKFOEINKOTKQMT 241 NQKLTLPENNSKVKULLQYSEASVALKKUNPKFOEINKOTKQMT

QY 301 YIWLWDISAISPKDQEISASRFEEDNEELRSRSTERHAWRNFTIV 301 YIWLWDISAISPKDQEISASRFEEDNEELRSRSTERHAWRNFTIV

US-10-120-801-88 ;  
TYPE: PRT ;  
ORGANISM: human ;  
SEQ ID NO: 88 ;  
LENGTH: 847 ;  
; ORGANISM: human ;  
US-10-120-801-88 ;

RESULT 16  
Sequence 88, Application US/10120801 ;  
Publication No. US2003020383A1 ;  
GENERAL INFORMATION:  
APPLICANT: Guo, Xiaojia  
APPLICANT: Pena, Carol  
APPLICANT: Shimkets, Richard  
APPLICANT: Radigaru, Murliidhara  
APPLICANT: Kekuda, Ramesh  
APPLICANT: Spytek, Kimberly  
APPLICANT: Mehraban, Fuad  
APPLICANT: Topper, James N  
APPLICANT: Nalyankar, Uriel  
APPLICANT: Wasserman, Scott  
APPLICANT: Edinger, Shlomit  
APPLICANT: Smitsion, Gleinda  
APPLICANT: Gunther, Erik  
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 21402-340  
CURRENT APPLICATION NUMBER: US/10/120,801  
CURRENT FILING DATE: 2002-04-11  
PRIOR APPLICATION NUMBER: 60/285748  
PRIOR FILING DATE: 2001-04-23  
PRIOR APPLICATION NUMBER: 60/286068  
PRIOR FILING DATE: 2001-04-24  
PRIOR APPLICATION NUMBER: 60/286292  
PRIOR FILING DATE: 2001-04-25  
PRIOR APPLICATION NUMBER: 60/288334  
PRIOR FILING DATE: 2001-05-03  
PRIOR APPLICATION NUMBER: 60/291241  
PRIOR FILING DATE: 2001-05-16  
PRIOR APPLICATION NUMBER: 60/322284  
PRIOR FILING DATE: 2001-09-14  
PRIOR APPLICATION NUMBER: 60/285609  
PRIOR FILING DATE: 2001-04-20  
NUMBER OF SEQ ID NOS: 155  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO: 88  
LENGTH: 847  
; TYPE: PRT  
; ORGANISM: human  
US-10-120-801-88 ;

Query Match 91.2%; Score 4473; DB 15; Length 845;  
Best Local Similarity 99.8%; Pred. No. 0; Mismatches 1; Indels 1;  
Matches 845; Conservative 1; MisMatches 1; Indels 1;

QY 1 MDFKLIQROTYTCVLSRHYGLYXCFGLGVVUTWSAOFGEVWLEWSRDOYI 61  
Db 1 MDFKLIQROTYTCVLSRHYGLYXCFGLGVVUTWSAOFGEVWLEWSRDOYI

QY 61 AKSFONRLCLPMLPIDVYTWTNGTDLLELKLQYREQMEEQKARRE 61  
Db 61 AKSFONRLCLPMLPIDVYTWTNGTDLLELKLQYREQMEEQKARRE

QY 121 KSFKQLECLLTCIKUMLVDPALPANTIKDPSLPSHSAISFNT 121  
Db 121 KSFKQLECLLTCIKUMLVDPALPANTIKDPSLPSHSAISFNT

QY 181 SVWVFSDSTKDVEDAHSGLKGSNSQTWRGMLTDKEVPGIVLWMDLAFT 181 SVWVFSDSTKDVEDAHSGLKGSNSQTWRGMLTDKEVPGIVLWMDLAFT

QY 241 NQKLTLPENNSKVKULLQYSEASVALKKUNPKFOEINKOTKQMT 241 NQKLTLPENNSKVKULLQYSEASVALKKUNPKFOEINKOTKQMT

QY 301 YIWLWDISAISPKDQEISASRFEEDNEELRSRSTERHAWRNFTIV 301 YIWLWDISAISPKDQEISASRFEEDNEELRSRSTERHAWRNFTIV

US-10-120-801-88 ;  
TYPE: PRT ;  
ORGANISM: human ;  
SEQ ID NO: 88 ;  
LENGTH: 847 ;  
; ORGANISM: human ;  
US-10-120-801-88 ;

RESULT 14  
US-10-120-801-88  
Sequence 88, Application US/10120801  
Publication No.US20030203843A1  
GENERAL INFORMATION:  
APPLICANT: Perna, Carol  
APPLICANT: Guo, Xiaoja  
APPLICANT: Shlimkers, Richard  
APPLICANT: Padigaru, Muralidhara  
APPLICANT: Kokuda, Ramesh  
APPLICANT: Soyer, Kimberly  
APPLICANT: Mehraban, Firday  
APPLICANT: Topper, James N.  
APPLICANT: Melnykier, Uriel  
APPLICANT: Wasserman, Scott  
APPLICANT: Edinger, Shlomit  
APPLICANT: Shulson, Glenda  
APPLICANT: Guntther, Erik  
APPLICANT: Komuves, Lazlo  
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 21402-340  
CURRENT APPLICATION NUMBER: US/10/120,801  
CURRENT FILING DATE: 2002-04-11  
PRIOR APPLICATION NUMBER: 60/285748  
PRIOR FILING DATE: 2001-04-23  
PRIOR APPLICATION NUMBER: 60/286068  
PRIOR FILING DATE: 2001-04-24  
PRIOR APPLICATION NUMBER: 60/286292  
PRIOR FILING DATE: 2001-04-25  
PRIOR APPLICATION NUMBER: 60/288334  
PRIOR FILING DATE: 2001-05-03  
PRIOR APPLICATION NUMBER: 60/291241  
PRIOR FILING DATE: 2001-05-16  
PRIOR APPLICATION NUMBER: 60/322284  
PRIOR FILING DATE: 2001-09-14  
PRIOR APPLICATION NUMBER: 60/285609  
PRIOR FILING DATE: 2001-04-20  
NUMBER OF SEQ ID NOS: 155  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 88  
LENGTH: 847  
TYPE: PRT  
ORGANISM: human  
US-10-120-801-88

QY 361 DNPRVTIVTHQDVRLNLSHLPTFSSPA1ESHIRIEGLSQKTYIYNDVMFCKDVMPDF 420  
Db 361 DNPRVTIVTHQDVRLNLSHLPTFSSPA1ESHIRIEGLSQKTYIYNDVMFCKDVMPDF 420  
QY 421 YSHRGQKQYLTWNPVNGAEGCPSWINDGYCDKACNNACWDG3DCCSGNSGSRYTAG 480  
Db 421 YSHRGQKQYLTWNPVNCAGCFCGSWIKDGYCDKACNNACWDG3DCCSGNSGSRYTAG 480  
QY 481 GGGTGSIGVGHQPGGGINSVSYCNOCANSFLAKFCDOACNVISCGFDAGDCQDFH 540  
Db 481 GGGTGSIGVGHQPGGGINSVSYCNOCANSFLAKFCDOACNVISCGFDAGDCQDFH 540  
QY 541 HELVKVILLPNOHTYI1PKGECLPYFSSPAEVAKRGVEGAYSNPLRHASTANKWTHL 600  
Db 541 HELVKVILLPNOHTYI1PKGECLPYFSSPAEVAKRGVEGAYSNPLRHASTANKWTHL 600  
QY 601 IMHSGMNAATHFLNTFONTBEPKQIIVTVEUDTRGPKLNSTAQKGYENLVSPTLLP 660  
Db 601 IMHSGMNAATHFLNTFONTBEPKQIIVTVEUDTRGPKLNSTAQKGYENLVSPTLLP 660  
QY 661 EAELFEDIPKECRPKPKRHDVNSTRQAEEVKRQVNLSPITLQEH 720  
Db 661 EAELFEDIPKECRPKPKRHDVNSTRQAEEVKRQVNLSPITLQEH 720  
QY 721 GDTTLKGMLSKSALLRSLMSQHAKRNQAITDETNSLVAPOBKQVHSILNSLG 780  
Db 721 GDTTLKGMLSKSALLRSLMSQHAKRNQAITDETNSLVAPOBKQVHSILNSLG 780  
QY 781 VSERLQRUTTPAVSKVNGHDOQNPPDLETTARFRVETHQTKTIGGNVTEKEPSSLIV 840  
Db 781 VSERLQRUTTPAVSKVNGHDOQNPPDLETTARFRVETHQTKTIGGNVTEKEPSSLIV 840  
QY 841 PLESQM 847  
Db 841 PLESQM 847  
  
RESULT 15  
Sequence 2392, Application US/10094749  
Publication No. US20030219741A1  
GENERAL INFORMATION:  
APPLICANT: ISOGAI, TAKAO  
APPLICANT: OGAWA, TOMOYASU  
APPLICANT: OTSUKI, TETSUJI  
APPLICANT: WAKAMATSU, AI  
APPLICANT: SATO, HIROYUKI  
APPLICANT: ISHII, SHIZUKO  
APPLICANT: YAMAMOTO, JUN-ICHI  
APPLICANT: ISONO, YUKO  
APPLICANT: HIO, YURI  
APPLICANT: OTSUKA, KAORU  
APPLICANT: NAGAI, KEIICHI  
APPLICANT: IRIE, RYOTARO  
APPLICANT: TAMEKIRI, ICHIRO  
APPLICANT: SEKI, NAOKO  
APPLICANT: YOSHIKAWA, TSUTOMU  
APPLICANT: OTSUKA, MOTOKI  
APPLICANT: NAGAHAMA, KENJI  
APPLICANT: MASUHO, YASUHIKO  
TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA  
FILE REFERENCE: 084335/0160  
CURRENT APPLICATION NUMBER: US/10/094,749  
CURRENT FILING DATE: 2002-03-12  
PRIOR APPLICATION NUMBER: 60/350,435  
PRIOR FILING DATE: 2002-01-24  
PRIOR APPLICATION NUMBER: JP 2001-328381  
PRIOR FILING DATE: 2001-09-14  
NUMBER OF SEQ ID NOS: 3381  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2392  
LENGTH: 846  
TYPE: PRT  
  
; ORGANISM: Homo sapiens  
; US-10-657-2392  
Query Match 91.1%; Score 4468; DB 15; Length 846;  
Best Local Similarity 99.8%; Pred. No. 0; Mismatches 1; Index 0; Gaps 0;  
Matches 844; Conservative 1; MisMatches 1; Indels 0; Gaps 0;  
  
QY 1 MLFKLILQRLQRTTCSSHRLQGLYVFLGVVUTSAFQGEVVLWSRQYHVLFDSDYDN 60  
Db 1 MLFKLILQRLQRTTCSSHRLQGLYVFLGVVUTSAFQGEVVLWSRQYHVLFDSDYDN 60  
QY 61 AGKSRFRNRCIPLPMIDVYUWTNGTDLKELQVREBOMEERQKAREBLIGKNTTEPK 120  
Db 61 AGKSRFRNRCIPLPMIDVYUWTNGTDLKELQVREBOMEERQKAREBLIGKNTTEPK 120  
QY 121 KSEKOLECLTHC1KVPMVLDPALPANITLKDOLPLSPHSASDIFNVAKEPNRSTNV 180  
Db 121 KSEKOLECLTHC1KVPMVLDPALPANITLKDOLPLSPHSASDIFNVAKEPNRSTNV 180  
QY 181 SVVTFDSTKVEDAHGLGLKNSRQTWGRYLTDEKEYVGLJLMQDAPLSGFPPTFET 240  
Db 181 SVVTFDSTKVEDAHGLGLKNSRQTWGRYLTDEKEYVGLJLMQDAPLSGFPPTFET 240  
QY 241 NOLKTKLPENLSSKKVLUQDYESASVALLKJNNPKDQBENLKQTKNTDGEKLTSPA 300  
Db 241 NOLKTKLPENLSSKKVLUQDYESASVALLKJNNPKDQBENLKQTKNTDGEKLTSPA 300  
QY 301 YLLMDLSAISQSOKDEDISASRFEDNEELRYSLSRISERHAPWNFIVINGQIPSNL 360  
Db 301 YLLMDLSAISQSOKDEDISASRFEDNEELRYSLSRISERHAPWNFIVINGQIPSNL 360  
QY 361 DNPRVTIVTHQDVRLNLSHLPTFSSPA1ESHIRIEGLSQKTYIYNDVMFCKDVMPDF 420  
Db 361 DNPRVTIVTHQDVRLNLSHLPTFSSPA1ESHIRIEGLSQKTYIYNDVMFCKDVMPDF 420  
Qy 421 YSHRGQKQYLTWNPVNCAGCFCGSWIKDGYCDKACNNACWDG3DCCSGNSGSRYTAG 480  
Db 421 YSHRGQKQYLTWNPVNCAGCFCGSWIKDGYCDKACNNACWDG3DCCSGNSGSRYTAG 480  
Qy 481 GGGTGSIGVGHQPGGGINSVSYCNOCANSFLAKFCDOACNVISCGFDAGDCQDFH 540  
Db 481 GGGTGSIGVGHQPGGGINSVSYCNOCANSFLAKFCDOACNVISCGFDAGDCQDFH 540  
Qy 541 HELVKVILLPNOHTYI1PKGECLPYFSSPAEVAKRGVEGAYSNPLRHASTANKWTHL 600  
Db 541 HELVKVILLPNOHTYI1PKGECLPYFSSPAEVAKRGVEGAYSNPLRHASTANKWTHL 600  
Qy 601 IMHSGMNAATHFLNTFONTBEPKQIIVTVEUDTRGPKLNSTAQKGYENLVSPTLLP 660  
Db 601 IMHSGMNAATHFLNTFONTBEPKQIIVTVEUDTRGPKLNSTAQKGYENLVSPTLLP 660  
Qy 661 EAELFEDIPKECRPKPKRHDVNSTRQAEEVKRQVNLSPITLQEH 720  
Db 661 EAELFEDIPKECRPKPKRHDVNSTRQAEEVKRQVNLSPITLQEH 720  
Qy 721 GDTTLKGMLSKSALLRSLMSQHAKRNQAITDETNSLVAPOBKQVHSILNSLG 780  
Db 721 GDTTLKGMLSKSALLRSLMSQHAKRNQAITDETNSLVAPOBKQVHSILNSLG 780  
Qy 781 VSERLQRUTTPAVSKVNGHDOQNPPDLETTARFRVETHQTKTIGGNVTEKEPSSLIV 840  
Db 781 VSERLQRUTTPAVSKVNGHDOQNPPDLETTARFRVETHQTKTIGGNVTEKEPSSLIV 840  
Qy 841 PLESQM 846  
Db 841 PLESQM 846

Search completed: July 26, 2004, 11:25:25  
Job time : 74.3389 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

Om protein - protein search, using sw model

Run on: July 26, 2004, 11:01:58 : Search time 15.4568 Seconds

(without alignments)

3126.212 Million cell updates/sec

Title: US-10-657-280-1

Perfect score: 4907

Sequence: 1 MFLKLGLQRTYTCIISHRVGL.....SLKTQIAYFTPSKNTQRQLK 928

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues  
Total number of hits satisfying chosen parameters: 141681  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	168.5	3.4	2470	1 NTC2_MOUSE
2	168.5	3.4	2471	1 NTC2_HUMAN
3	168.5	3.4	2471	1 NTC2_RAT
4	166.5	3.4	1429	1 L112_CABEL
5	165.5	3.4	2531	1 NTC1_RAT
6	163	3.3	2531	1 NTC1_MOUSE
7	162.5	3.3	2437	1 NTC1_BRARE
8	159.5	3.1	2321	1 NTC3_HUMAN
9	152	3.1	1964	1 NTC4_MOUSE
10	151.5	3.1	2003	1 NTC4_HUMAN
11	149.5	3.0	2524	1 NOTC_XENLA
12	148	3.0	2319	1 NTC3_RAT
13	147.5	3.0	2318	1 NTC3_MOUSE
14	145.5	3.0	2358	1 NTC1_HUMAN
15	145	3.0	626	1 GLP1_BUGBP
16	137.5	2.8	1328	1 YM00_YEAST
17	136.5	2.8	1701	1 MSPI_PLAFW
18	135.5	2.8	1701	1 MSPI_PLAFF
19	134.5	2.7	2324	1 YCF1_OENHO
20	134	2.7	1295	1 GLP1_CABEL
21	131	2.7	2095	1 RPL_MOUSE
22	129.5	2.6	1630	1 MST1_PLAFK
23	129.5	2.6	1639	1 MSPI_PLAFW
24	129.5	2.6	2867	1 RBP2_PLAFW
25	129	2.6	2703	1 NOTC_DROME
26	128.5	2.6	756	1 Y328_MYGB
27	128	2.6	631	1 GLD_BUCA
28	127	2.6	780	1 MUSS_BOBO
29	127	2.6	1459	1 GEA2_YEAST
30	126.5	2.6	1002	1 CLMN_HUMAN
31	126	2.6	1328	1 YM75_YEAST
32	125.5	2.6	2778	1 PGCV_RAT
33	124.5	2.5	1 S230_PLAFO	

#### ALIGNMENTS

34	124	2.5	3259	1 GO31_HUMAN	Q14789 homo sapien
35	123	2.5	770	1 DB94_YEAST	P20448 saccharomyces
36	123	2.5	5420	1 MACF_HUMAN	Q99ab4 homo sapien
37	122.5	2.5	111	1 EX52_BUFBP	P23201 saccharomyces
38	122.5	2.5	1466	1 SP42_YEAST	P38873 saccharomyces
39	122.5	2.5	1541	1 YH6_YEAST	P09278 varicella-zoster virus
40	122.5	2.5	2763	1 TEGU_VIZD	P12753 saccharomyces
41	122	2.5	1312	1 RAS0_YEAST	Q96pk2 homo sapien
42	122	2.5	5938	1 MAC4_HUMAN	P47100 saccharomyces
43	121.5	2.5	1755	1 YU29_YEAST	P53804 homo sapien
44	121.5	2.5	2052	1 TTC3_HUMAN	Q8c5w0 mus musculus
45	121	2.5	1052	1 CLMN_MOUSE	

RT proteolysis.";  
 J. Biol. Chem. 276:40268-40273 (2001).  
 R.L. RN [7].

RT POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.  
 RX MEDLINE-213>4376; Pubmed=1459941; RA Mizutani T., Taniguchi Y., Aki T., Hashimoto N., Honjo T.; RT "Conservation of the biochemical mechanisms of signal transduction among mammalian Notch family members";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031 (2001).  
 CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.  
 CC Upon ligand activation through the released notch intracellular domain (NICD) it forms a transcriptional activator complex with  
 CC RBP-J kappa and activates genes of the enhancer of split locus.  
 CC Affects the implementation of differentiation, proliferation and  
 CC apoptotic programs (By similarity). May play an essential role in  
 CC postimplantation development, probably in some aspect of cell  
 CC specification and/or differentiation.  
 CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-  
 CC terminal fragment N(EC) which are probably linked by disulfide  
 bonds.  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following  
 CC proteolytic processing NICD is translocated to the nucleus.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1; IsoId=035516-1; Sequence=Displayed;

Name=2;  
 CC IsoId=035516-2; Sequence=vSP\_001405;  
 CC Note=No experimental confirmation available;  
 CC TRISUB SPECIFICITY: Expressed in the brain, liver, kidney,  
 CC neuroepithelia, somites, optic vesicles and branchial arches, but  
 CC not heart.  
 CC -!- DEVELOPMENTAL STAGE: Expressed in the embryonic ventricular zone,  
 CC the postnatal ependymal cells, and the choroid plexus throughout  
 CC embryonic and postnatal development.  
 CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form  
 CC which is proteolytically cleaved by a furin-like convertase in the  
 CC trans-Golgi network before it reaches the plasma membrane to yield  
 CC an active, ligand-accessible form. Cleavage results in a C-  
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following  
 CC lisinopril binding, it is cleaved by TNF-alpha converting enzyme  
 CC (TACE) to yield a membrane-associated intermediate fragment called  
 CC notch extracellular truncation (NEXT). This fragment is then  
 CC cleaved by presenilin dependent gamma-secretase to release a  
 CC notch-derived peptide containing the intracellular domain (NICD)  
 CC from the membrane.  
 CC -!- PTM: Phosphorylated.  
 CC -!- SIMILARITY: Belongs to the NOTCH family.  
 CC -!- SIMILARITY: Contains 35 EGF-like domains.  
 CC -!- SIMILARITY: Contains 2 Lin/Notch repeats.  
 CC -!- SIMILARITY: Contains 5 ANK repeats.

---

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR InterPro: IPRO00742; EGF-2.  
 DR InterPro: IPRO01891; EGF-Ca.  
 DR InterPro: IPRO1438; EGF-II.  
 DR InterPro: IPRO06209; EGF-like.  
 DR InterPro: IPRO009; Laminin\_EGF.  
 DR InterPro: IPRO08297; Notch.  
 DR InterPro: IPRO00800; Notch\_dom.  
 DR Pfam: PF00023; ank; 6.  
 DR Pfam: PF0008; EGF; 34.  
 DR SMART; SNO019; notch; 2.  
 SMART; SNO0094; NL; 3.  
 DR PIRSP; PIRSF00279; Notch; 1.  
 DR PROSITE; PS50297; ANK\_REPEAT; 4.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 22.  
 DR PROSITE; PS00022; EGF\_1; 33.  
 DR PROSITE; PS0116; EGF\_2; 27.  
 DR PROSITE; PS00026; EGF\_3; 35.  
 DR Receptor; Transcription regulation; Activator; Differentiation; Developmental protein; Repat; ANK repeat; EGF-like domain; Transmembrane; Glycoprotein; Signal; Phosphorylation; KW Alternative splicing.  
 FT SIGNAL; 1; 25.  
 FT CHAIN; 26; 2470.  
 FT CHAIN; 1666; 2470.  
 FT CHAIN; 1679; 2470.  
 FT DOMAIN; 26; 1677.  
 FT TRANSMEM; 1678; 1698.  
 FT DOMAIN; 1679; 2470.  
 FT DOMAIN; 26; 63.  
 FT DOMAIN; 64; 102.  
 FT DOMAIN; 105; 143.  
 FT DOMAIN; 144; 180.  
 FT DOMAIN; 182; 219.  
 FT DOMAIN; 221; 256.  
 FT DOMAIN; 258; 294.  
 FT DOMAIN; 266; 334.  
 FT DOMAIN; 336; 372.  
 FT DOMAIN; 373; 411.  
 FT DOMAIN; 413; 452.  
 FT DOMAIN; 450; 490.  
 FT DOMAIN; 492; 528.  
 FT DOMAIN; 530; 566.  
 FT DOMAIN; 568; 603.  
 FT DOMAIN; 605; 641.  
 FT DOMAIN; 643; 678.  
 FT DOMAIN; 680; 716.  
 FT DOMAIN; 718; 753.  
 FT DOMAIN; 755; 791.  
 FT DOMAIN; 793; 829.  
 FT DOMAIN; 831; 869.  
 FT DOMAIN; 871; 907.  
 FT DOMAIN; 909; 945.  
 FT DOMAIN; 947; 983.  
 FT DOMAIN; 1021; 1059.  
 FT DOMAIN; 1023; 1059.  
 FT DOMAIN; 1061; 1097.  
 FT DOMAIN; 1099; 1145.  
 FT DOMAIN; 1147; 1183.  
 FT DOMAIN; 1185; 1221.  
 FT DOMAIN; 1223; 1260.  
 FT DOMAIN; 1262; 1300.  
 FT DOMAIN; 1304; 1345.  
 FT DOMAIN; 1372; 1410.  
 REPEAT; 1418; 1454.  
 REPEAT; 1501; 1539.  
 REPEAT; 1829

DR InterPro: IPRO00742; EGF-2.  
 DR InterPro: IPRO01891; EGF-Ca.  
 DR InterPro: IPRO1438; EGF-II.  
 DR InterPro: IPRO06209; EGF-like.  
 DR InterPro: IPRO009; Laminin\_EGF.  
 DR InterPro: IPRO08297; Notch.  
 DR InterPro: IPRO00800; Notch\_dom.  
 DR Pfam: PF00023; ank; 6.  
 DR Pfam: PF0008; EGF; 34.  
 DR SMART; SNO019; notch; 2.  
 SMART; SNO0094; NL; 3.  
 DR PIRSP; PIRSF00279; Notch; 1.  
 DR PROSITE; PS50297; ANK\_REPEAT; 4.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 22.  
 DR PROSITE; PS00022; EGF\_1; 33.  
 DR PROSITE; PS0116; EGF\_2; 27.  
 DR PROSITE; PS00026; EGF\_3; 35.  
 DR Receptor; Transcription regulation; Activator; Differentiation; Developmental protein; Repat; ANK repeat; EGF-like domain; Transmembrane; Glycoprotein; Signal; Phosphorylation; KW Alternative splicing.  
 FT SIGNAL; 1; 25.  
 FT CHAIN; 26; 2470.  
 FT CHAIN; 1666; 2470.  
 FT CHAIN; 1679; 2470.  
 FT DOMAIN; 26; 1677.  
 FT TRANSMEM; 1678; 1698.  
 FT DOMAIN; 1679; 2470.  
 FT DOMAIN; 26; 63.  
 FT DOMAIN; 64; 102.  
 FT DOMAIN; 105; 143.  
 FT DOMAIN; 144; 180.  
 FT DOMAIN; 182; 219.  
 FT DOMAIN; 221; 256.  
 FT DOMAIN; 258; 294.  
 FT DOMAIN; 266; 334.  
 FT DOMAIN; 336; 372.  
 FT DOMAIN; 373; 411.  
 FT DOMAIN; 413; 452.  
 FT DOMAIN; 450; 490.  
 FT DOMAIN; 492; 528.  
 FT DOMAIN; 530; 566.  
 FT DOMAIN; 568; 603.  
 FT DOMAIN; 605; 641.  
 FT DOMAIN; 643; 678.  
 FT DOMAIN; 680; 716.  
 FT DOMAIN; 718; 753.  
 FT DOMAIN; 755; 791.  
 FT DOMAIN; 793; 829.  
 FT DOMAIN; 831; 869.  
 FT DOMAIN; 871; 907.  
 FT DOMAIN; 909; 945.  
 FT DOMAIN; 947; 983.  
 FT DOMAIN; 1021; 1059.  
 FT DOMAIN; 1023; 1059.  
 FT DOMAIN; 1061; 1097.  
 FT DOMAIN; 1099; 1145.  
 FT DOMAIN; 1147; 1183.  
 FT DOMAIN; 1185; 1221.  
 FT DOMAIN; 1223; 1260.  
 FT DOMAIN; 1262; 1300.  
 FT DOMAIN; 1304; 1345.  
 FT DOMAIN; 1372; 1410.  
 REPEAT; 1418; 1454.  
 REPEAT; 1501; 1539.  
 REPEAT; 1829



KW	Transmembrane; Glycoprotein; Signal; Phosphorylation.	FT	DISULFID	248	257	BY SIMILARITY.
PT	SIGNAL	1	25			
PT	CHAIN	26	2471			
PT	CHAIN	1666	2471			
PT	CHAIN	1697	2471			
PT	DOMAIN	26	1677			
PT	TRANSMEM	1678	1698			
PT	DOMAIN	1699	2471			
PT	DOMAIN	26	63			
PT	DOMAIN	64	102			
PT	DOMAIN	105	143			
PT	DOMAIN	144	180			
PT	DOMAIN	182	219			
PT	DOMAIN	221	258			
PT	DOMAIN	260	296			
PT	DOMAIN	298	336			
PT	DOMAIN	338	374			
PT	DOMAIN	375	413			
PT	DOMAIN	415	454			
PT	DOMAIN	456	492			
PT	DOMAIN	494	530			
PT	DOMAIN	532	568			
PT	DOMAIN	570	605			
PT	DOMAIN	607	643			
PT	DOMAIN	645	680			
PT	DOMAIN	682	718			
PT	DOMAIN	720	755			
PT	DOMAIN	757	793			
PT	DOMAIN	795	831			
PT	DOMAIN	833	871			
PT	DOMAIN	873	909			
PT	DOMAIN	911	947			
PT	DOMAIN	949	985			
PT	DOMAIN	987	1023			
PT	DOMAIN	1025	1061			
PT	DOMAIN	1063	1099			
PT	DOMAIN	1101	1147			
PT	DOMAIN	1149	1185			
PT	DOMAIN	1187	1223			
PT	DOMAIN	1225	1262			
PT	DOMAIN	1264	1302			
PT	DOMAIN	1304	1343			
PT	DOMAIN	1374	1412			
PT	REPEAT	1420	1456			
PT	REPEAT	1450	1503			
PT	REPEAT	1503	1535			
PT	REPEAT	1535	1571			
PT	REPEAT	1571	1871			
PT	REPEAT	1871	ANK 1.			
PT	REPEAT	1875	ANK 2.			
PT	REPEAT	1905	1939			
PT	REPEAT	1939	1972			
PT	REPEAT	1943	1972			
PT	REPEAT	1976	2005			
PT	REPEAT	2009	2038			
PT	REPEAT	2038	2058			
PT	REPEAT	2058	2095			
PT	REPEAT	2095	2134			
PT	REPEAT	2134	2173			
PT	REPEAT	2173	2212			
PT	REPEAT	2212	2251			
PT	REPEAT	2251	2289			
PT	REPEAT	2289	2429			
PT	REPEAT	2429	2429			
PT	DISULFID	28	41			
PT	DISULFID	35	51			
PT	DISULFID	53	62			
PT	DISULFID	68	79			
PT	DISULFID	73	90			
PT	DISULFID	92	101			
PT	DISULFID	109	121			
PT	DISULFID	115	131			
PT	DISULFID	133	142			
PT	DISULFID	148	159			
PT	DISULFID	153	168			
PT	DISULFID	170	179			
PT	DISULFID	186	198			
PT	DISULFID	192	207			
PT	DISULFID	209	218			
PT	DISULFID	225	236			
PT	DISULFID	230	246			

  

FT	DISULFID	248	257	BY SIMILARITY.
Query Match	3 4%;	Score	168.5;	DB 1; Length 2471;
Best Local Similarity	19.7%;	Pred.	No. 042;	
Matches	114;	Indels	215;	Gaps 27;
Conservative	66;	Mismatches	185;	
NM				
Similarity				
NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 2.				
NOTCH EXTRACELLULAR TRUNCATION (BY SIMILARITY).				
NOTCH INTRACELLULAR DOMAIN (BY SIMILARITY).				
EXTRACELLULAR (POTENTIAL).				
CYTOPLASMIC (POTENTIAL).				
EGF-LIKE 1.				
EGF-LIKE 2.				
EGF-LIKE 3.				
EGF-LIKE 4.				
EGF-LIKE 5.				
EGF-LIKE 6.				
EGF-LIKE 7.				
EGF-LIKE 8.				
EGF-LIKE 9.				
EGF-LIKE 10.				
EGF-LIKE 11.				
EGF-LIKE 12.				
EGF-LIKE 13.				
EGF-LIKE 14.				
EGF-LIKE 15.				
EGF-LIKE 16.				
EGF-LIKE 17.				
EGF-LIKE 18.				
EGF-LIKE 19.				
EGF-LIKE 20.				
EGF-LIKE 21.				
EGF-LIKE 22.				
EGF-LIKE 23.				
EGF-LIKE 24.				
EGF-LIKE 25.				
EGF-LIKE 26.				
EGF-LIKE 27.				
EGF-LIKE 28.				
EGF-LIKE 29.				
LIN/NOTCH 1.				
LIN/NOTCH 2.				

  

RESULT	3
NTC2_RAT	NTC2_RAT
ID	NTC2_RAT
AC	Q80K30;
DT	28-FEB-2003 (Rel. 41, Last sequence update)
DT	28-FEB-2003 (Rel. 41, last annotation update)
DE	Neurogenic locus notch homolog protein 2 precursor (Notch 2).
GN	NOTCH2.
OS	Rattus norvegicus (Rat).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX	NCBI_TaxID=10116;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE:Brain;
RX	MEIDLINE-9320075; Pubmed-1295745;
RA	Weinmaster G., Roberts V.J., Lemke G.;
RT	"Notch2: a second mammalian Notch gene.";
RL	Development 116:931-941(1992).
RN	[2]
TP	TISSUE:SPECIFICITY.
RA	Irwin D.K., Zurich S.D., Nguyen T., Weinmaster G., Komblum H.I.;
RT	"Expression patterns of Notch1, Notch2, and Notch3 suggest multiple functional roles for the Notch-DSL signaling system during brain

RT	PROS1; PS00022; EGF-1;	34.
RL	PROS1; PS00186; EGF-2;	26.
CC	PROS1; PS0026; EGF-3;	35.
CC	DR PROS1; PS00177; EGF-CA;	22.
CC	KW Receptor; Transcription regulation; Activator; Differentiation;	
CC	KW Developmental protein; Repeat; EGF-like domain;	
CC	KW Transmembrane; Glycoprotein; Signal; Phosphorylation.	
AFFECTS THE IMPLEMENTATION OF DIFFERENTIATION, PROLIFERATION AND APOPTOTIC PROGRAMS. MAY PLAY AN ESSENTIAL ROLE IN POSTIMPLANTATION DEVELOPMENT, PROBABLY IN SOME ASPECT OF CELL SPECIFICATION AND/OR DIFFERENTIATION. (BY SIMILARITY).	PT SIGNAL 1	25
-!- SUBUNIT: Heterodimer of a C-terminal fragment N(EGF) which are probably linked by disulfide bonds (By similarity).	PT CHAIN 26	2471
-!- SUBCELLULAR LOCATION: Type I membrane protein. Following proteolytical processing NICD is translocated to the nucleus.	PT DOMAIN 26	1677
-!- TISSUE SPECIFICITY: Highly expressed in the spleen and choroid plexus in the brain. Expressed in postnatal central nervous system (CNS) germinal zones and, in early postnatal life, within numerous cells throughout the CNS. It is more highly localized to ventricular germinal zones. Also found in the heart, liver and kidney.	PT DOMAIN 26	1698
-!- DEVELOPMENTAL STAGE: Expressed in the brain during E14 and E17.	PT DOMAIN 26	1699
-!- PTM: Synthesized in the endoplasmic reticulum as an inactive form which is proteolytically cleaved by a furin-like convertase in the trans-Golgi network before it reaches the plasma membrane to yield an active ligand-accessible form. Cleavage results in a C-terminal fragment N(EGF) and a N-terminal fragment N(EGF). Following ligand binding, it is cleaved by TNF-alpha converting enzyme (TACE) to yield a membrane-associated intermediate fragment called notch extracellular truncation (NEXT). This fragment is then cleaved by presenilin dependent gamma-secretase to release a notch-derived peptide containing the intracellular domain (NICD) from the membrane (By similarity).	PT DOMAIN 26	2471
-!- PTM: Phosphorylated (By similarity).	PT DOMAIN 26	63
-!- SIMILARITY: Belongs to the NOTCH family.	PT DOMAIN 64	102
-!- SIMILARITY: Contains 35 EGF-like domains.	PT DOMAIN 105	143
-!- SIMILARITY: Contains 2 Lin/Notch repeats.	PT DOMAIN 144	180
-!- SIMILARITY: Contains 6 ANK repeats.	PT DOMAIN 182	219
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@isb-sib.ch).	FT DOMAIN 221	258
EMBL; M93651; AAK13588.1; -.	FT DOMAIN 260	295
PIR; A49128.	FT DOMAIN 298	336
ASSP; P00743; ICCC-1.	FT DOMAIN 338	374
DR Interpro; IPR002110; ANK.	FT DOMAIN 375	413
DR Interpro; IPR00080; Notch.	FT DOMAIN 415	454
DR Interpro; IPR00742; EGF-2.	FT DOMAIN 456	492
DR Interpro; IPR001881; EGF-Ca.	FT DOMAIN 494	530
DR Interpro; IPR001438; EGF-II.	FT DOMAIN 532	568
DR Interpro; IPR006209; EGF-Like.	FT DOMAIN 570	605
DR Interpro; IPR000249; Laminin-EGF.	FT DOMAIN 607	643
DR Interpro; IPR008297; Notch.	FT DOMAIN 645	680
DR Interpro; IPR000152; ASX hydroxyl_S.	FT DOMAIN 682	718
DR Interpro; IPR000741; EGF-1.	FT DOMAIN 720	755
DR Interpro; IPR001881; EGF-Ca.	FT DOMAIN 757	793
DR Interpro; IPR001438; EGF-II.	FT DOMAIN 795	831
DR Interpro; IPR006209; EGF-Like.	FT DOMAIN 833	871
DR Interpro; IPR000249; Laminin-EGF.	FT DOMAIN 873	909
DR Interpro; IPR008297; Notch.	FT DOMAIN 911	947
DR Interpro; IPR000742; EGF-2.	FT DOMAIN 949	985
DR Interpro; IPR001881; EGF-Ca.	FT DOMAIN 987	1023
DR Interpro; IPR001438; EGF-II.	FT DOMAIN 1025	1061
DR Interpro; IPR006209; EGF-Like.	FT DOMAIN 1063	1099
DR Interpro; IPR000249; Laminin-EGF.	FT DOMAIN 1101	1147
DR Interpro; IPR008297; Notch.	FT DOMAIN 1149	1185
DR Interpro; IPR000742; EGF-2.	FT DOMAIN 1187	1223
DR Interpro; IPR001881; EGF-Ca.	FT DOMAIN 1225	1262
DR Interpro; IPR001438; EGF-II.	FT DOMAIN 1264	1302
DR Interpro; IPR006209; EGF-Like.	FT DOMAIN 1304	1343
DR Interpro; IPR000249; Laminin-EGF.	FT DOMAIN 1374	1412
DR Interpro; IPR008297; Notch.	FT DOMAIN 1645	1648
DR Interpro; IPR000742; EGF-2.	FT DOMAIN 1994	1997
DR Interpro; IPR001881; EGF-Ca.	FT DOMAIN 2426	2429
DR Interpro; IPR001438; EGF-II.	FT DOMAIN 246	2451
DR Interpro; IPR006209; EGF-Like.	FT DOMAIN 1420	1456
DR Interpro; IPR000249; Laminin-EGF.	FT DOMAIN 1503	1535
DR Interpro; IPR008297; Notch.	FT DOMAIN 1827	1871
DR Interpro; IPR000742; EGF-2.	FT REPEAT 1876	1905
DR Interpro; IPR001881; EGF-Ca.	FT REPEAT 1909	1939
DR Interpro; IPR001438; EGF-II.	FT REPEAT 1943	1972
DR Interpro; IPR006209; EGF-Like.	FT REPEAT 1976	2005
DR Interpro; IPR000249; Laminin-EGF.	FT REPEAT 2009	2038
DR Interpro; IPR008297; Notch.	FT DISULFID 28	41
DR Interpro; IPR000742; EGF-2.	FT DISULFID 35	51
DR Interpro; IPR001881; EGF-Ca.	FT DISULFID 62	62
DR Interpro; IPR001438; EGF-II.	FT DISULFID 68	79
DR Interpro; IPR006209; EGF-Like.	FT DISULFID 73	90
DR Interpro; IPR000249; Laminin-EGF.	FT DISULFID 92	101
DR Interpro; IPR008297; Notch.	FT DISULFID 109	121
DR Interpro; IPR000742; EGF-2.	FT DISULFID 115	131
DR Interpro; IPR001881; EGF-Ca.	FT DISULFID 133	142
DR Interpro; IPR001438; EGF-II.	FT DISULFID 148	159

DISULFID 153 BY SIMILARITY.  
 FT DISULFID 170 BY SIMILARITY.  
 PT DISULFID 179 BY SIMILARITY.  
 FT DISULFID 186 BY SIMILARITY.  
 FT DISULFID 192 BY SIMILARITY.  
 FT DISULFID 209 BY SIMILARITY.  
 FT DISULFID 225 BY SIMILARITY.  
 FT DISULFID 230 BY SIMILARITY.  
 FT DISULFID 248 BY SIMILARITY.  
 FT DISULFID 264 BY SIMILARITY.  
 FT DISULFID 269 BY SIMILARITY.  
 FT DISULFID 286 BY SIMILARITY.  
 FT DISULFID 302 BY SIMILARITY.  
 FT DISULFID 309 BY SIMILARITY.  
 FT DISULFID 326 BY SIMILARITY.  
 FT DISULFID 342 BY SIMILARITY.  
 FT DISULFID 347 BY SIMILARITY.  
 FT DISULFID 364 BY SIMILARITY.  
 FT DISULFID 379 BY SIMILARITY.  
 FT DISULFID 384 BY SIMILARITY.  
 FT DISULFID 403 BY SIMILARITY.  
 FT DISULFID 412 BY SIMILARITY.  
 FT DISULFID 427 BY SIMILARITY.  
 FT DISULFID 444 BY SIMILARITY.  
 FT DISULFID 455 BY SIMILARITY.  
 FT DISULFID 482 BY SIMILARITY.  
 FT DISULFID 498 BY SIMILARITY.  
 FT DISULFID 503 BY SIMILARITY.  
 FT DISULFID 520 BY SIMILARITY.  
 FT DISULFID 536 BY SIMILARITY.  
  
 Query Match 3.4%; Score 168.5; DB 1; Length 2471;  
 Best Local Similarity 26.2%; Pred. No. 0\_042; Gaps 5;  
 Matches 44; Conservative 15; Mismatches 58; Indels 51; Gaps 5;  
  
 QY 417 PDDFYSHSKGQKYLITWVPVNGAEGCGESWIKGYDKACINSACTCWDGGDSGNSSGR 476  
 Db 1404 PPFWGSRSYTAAPTSPATPQSCYQCADKRGIDAECHNSHACQWDGGDC-----AN 1455  
 QY 477 YIAGGGTGSIGVQPHQFQGGGINVSYVCYNGC-----AN 511  
 Db 1456 -----SLTMEDPWANCTSSIRCYHINYNNQCDLNTAACLENFECORNISKCKYD 1506  
 QY 512 SWLADKPF---CDOACNTLSCFGDAAGRCQGDHFHELYK----VIIIP 550  
 Db 1507 KYCADHFKDNCDKGCNNEECGWDGLDCAALDPENLAEGILIVVILP 1554  
  
 RESULT 4  
 L112 CABEL STANDARD; PRT; 1429 AA.  
 ID L112 CABEL STANDARD; PRT; 1429 AA.  
 AC P1455;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE LIN-12 protein precursor.  
 GN LIN-12 OR R107.8.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID:6239;  
 RN [1] SEQUENCE FROM N.A.  
 RP STRAIN=Bristol N2;  
 RX MEDLINE=88334747; PubMed=3419531;  
 RA Yochim J., Weston K., Greenwald I.;  
 RT The *Caenorhabditis elegans* lin-12 gene encodes a transmembrane protein with overall similarity to *Drosophila* Notch. ";  
 Nature 335:547-550(1988).  
 [2] SEQUENCE FROM N.A.  
 RP STRAIN=Bristol N2;  
 MEDLINE=94150718; PubMed=7906398;

Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,' Berriman M.,' Bonsfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,' Craxton M., Dear S., Du Z., Dutkin R., Favell A., Fraser A., Fulton L., Gardner A., Green P., Hawkins J.,' Kershaw J.,' Kirsten N.,' Johnston L., Jones M.,' Kershaw J.,' Kirsten N.,' Latreille P.,' Lightning J.,' Lloyd C.,' Mortimore B.,' O'Callaghan M., Parsons J.,' Percy C.,' Rifkin L.,' Roopra A.,' Saunders D.,' Showkeen R., Sims M.,' Smalldon N.,' Smith A.,' Smith M.,' Sonnhammer E.,' Staden R.,' Sulston J.,' Thierry-Mieg J.,' Thomas K.,' Vaudin M.,' Vaughan K.,' Waterston R.,' Watson A.,' Weinstock L.,' Wilkinson-Spratt J.,' Wohldman P.;' "2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans";' Nature 368:33-38(1994).  
 RL -!- FUNCTION: Involved in several cell fates decisions that requires cell-cell interactions. It is possible that Lin-12 encodes a membrane-bound receptor for a signal that enables expression of the ventral uterine precursor cell fate.  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -!- SIMILARITY: HIGH, TO C\_ELEGANS\_GIP-1.  
 CC -!- SIMILARITY: Contains 13 EGF-like domains.  
 CC -!- SIMILARITY: Contains 3 Lin/Notch repeats.  
  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).  
 CC DR EMBL; M12069; AAA70191.1; -.  
 DR EMBL; Z14092; CAJ78414.1; -.  
 DR PIR; S06434; S06434.  
 DR HSSE; P00740; IEDB.  
 DR WormPep; R107.8; CE00374.  
 DR InterPro; IPR002110; ANK.  
 DR InterPro; IPR00152; AX\_hydroxyl\_S.  
 DR InterPro; IPR000742; EGF\_2.  
 DR InterPro; IPR001881; EGF\_Ca.  
 DR InterPro; IPR006309; EGF\_like.  
 DR InterPro; IPR002049; Notch\_EGF.  
 DR InterPro; IPR000800; Notch\_dom.  
 DR Pfam; PF00033; ank; 6.  
 DR Pfam; PF00008; EGF; 13.  
 DR Pfam; PF00066; notch; 3.  
 DR PRINTS; PRO0011; EGFLAMININ.  
 DR PRINTS; PRO452; NOTCH.  
 DR SMART; SM00248; ANK; 6.  
 DR SMART; SM00179; EGF\_Ca; 2.  
 DR SMART; SM0005; NL\_2.  
 DR PROSITE; PSS0297; ANK\_REPEAT\_REGION; 1.  
 DR PROSITE; PSS0088; ANK\_REPEAT; 3.  
 DR PROSITE; PS0010; ASY\_HYDROXYL; 3.  
 DR PROSITE; PS0010; ASY\_HYDROXYL; 3.  
 DR PROSITE; PS0022; EGF\_I; 12.  
 DR PROSITE; PS0186; EGF\_2; 11.  
 DR PROSITE; PS0026; EGF\_3; 13.  
 DR PROSITE; PS0187; EGF\_Ca; 2.  
 DR KW Differentiation; Repeat; ANK repeat; EGF-like domain; Transmembrane; Glycoprotein; Signal; KW SIGNAL; 1; 15; POTENTIAL.  
 FT CHAIN 16; 1429; LIN-12 PROTEIN.  
 FT DOMAIN 16; 1429; EXTRACHLORULAR (POTENTIAL).  
 FT TRANSMEM 909; 931; POTENTIAL.  
 FT DOMAIN 932; 1429; CYTOSPLASMIC (POTENTIAL).  
 FT DOMAIN 20; 61; EGF-LIKE 1.  
 FT DOMAIN 114; 150; EGF-LIKE 2.  
 FT DOMAIN 152; 190; EGF-LIKE 3.  
 FT DOMAIN 201; 246; EGF-LIKE 4.  
 FT DOMAIN 250; 285; EGF-LIKE 5.  
 FT DOMAIN 287; 323; EGF-LIKE 6.  
 FT DOMAIN 323; 363; EGF-LIKE 7.

FT DOMAIN	365	402	EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).	Qy 529 GFDAGDCGQDFHFLYKVILLPNOTHYIIPKGCLPLVFSFAVKRGVEGAVSDNPIRH 588
FT DOMAIN	404	441	EGF-LIKE 9.	RN
FT DOMAIN	449	492	EGF-LIKE 10.	RN
FT DOMAIN	503	541	EGF-LIKE 11.	RN
FT DOMAIN	543	579	EGF-LIKE 12.	RN
FT DOMAIN	582	619	EGF-LIKE 13.	RN
FT REPEAT	635	669	LIN/NOTCH 1.	RN
FT REPEAT	670	710	LIN/NOTCH 2.	RN
FT REPEAT	711	750	LIN/NOTCH 3.	RN
FT REPEAT	1093	1122	ANK 1.	RN
FT REPEAT	1126	1158	ANK 2.	RN
FT REPEAT	1162	1194	ANK 3.	RN
FT REPEAT	1205	1235	ANK 4.	RN
FT REPEAT	1240	1269	ANK 5.	RN
FT DISULFID	24	35	BY SIMILARITY.	RN
FT DISULFID	29	49	BY SIMILARITY.	RN
FT DISULFID	51	60	BY SIMILARITY.	RN
FT DISULFID	60	60	BY SIMILARITY.	RN
FT DISULFID	118	129	BY SIMILARITY.	RN
FT DISULFID	123	138	BY SIMILARITY.	RN
FT DISULFID	140	149	BY SIMILARITY.	RN
FT DISULFID	156	169	BY SIMILARITY.	RN
FT DISULFID	163	178	BY SIMILARITY.	RN
FT DISULFID	180	189	BY SIMILARITY.	RN
FT DISULFID	205	227	BY SIMILARITY.	RN
FT DISULFID	221	234	BY SIMILARITY.	RN
FT DISULFID	236	245	BY SIMILARITY.	RN
FT DISULFID	254	264	BY SIMILARITY.	RN
FT DISULFID	259	273	BY SIMILARITY.	RN
FT DISULFID	275	284	BY SIMILARITY.	RN
FT DISULFID	291	302	BY SIMILARITY.	RN
FT DISULFID	296	311	BY SIMILARITY.	RN
FT DISULFID	313	322	BY SIMILARITY.	RN
FT DISULFID	327	339	BY SIMILARITY.	RN
FT DISULFID	334	351	BY SIMILARITY.	RN
FT DISULFID	353	362	BY SIMILARITY.	RN
FT DISULFID	369	381	BY SIMILARITY.	RN
FT DISULFID	375	390	BY SIMILARITY.	RN
FT DISULFID	392	401	BY SIMILARITY.	RN
FT DISULFID	408	419	BY SIMILARITY.	RN
FT DISULFID	413	429	BY SIMILARITY.	RN
FT DISULFID	431	440	BY SIMILARITY.	RN
FT DISULFID	507	518	BY SIMILARITY.	RN
FT DISULFID	512	529	BY SIMILARITY.	RN
FT DISULFID	531	540	BY SIMILARITY.	RN
FT DISULFID	547	558	BY SIMILARITY.	RN
FT DISULFID	552	567	BY SIMILARITY.	RN
FT DISULFID	569	578	BY SIMILARITY.	RN
FT DISULFID	586	597	BY SIMILARITY.	RN
FT DISULFID	591	607	BY SIMILARITY.	RN
FT DISULFID	609	618	BY SIMILARITY.	RN
PT CARBOHYD	41	41	N-LINKED (GLCNAC. . ) (POTENTIAL).	RN
PT CARBOHYD	155	155	N-LINKED (GLCNAC. . ) (POTENTIAL).	RN
PT CARBOHYD	194	194	N-LINKED (GLCNAC. . ) (POTENTIAL).	RN
PT CARBOHYD	378	378	N-LINKED (GLCNAC. . ) (POTENTIAL).	RN
PT CARBOHYD	515	515	N-LINKED (GLCNAC. . ) (POTENTIAL).	RN
PT CARBOHYD	623	623	N-LINKED (GLCNAC. . ) (POTENTIAL).	RN
PT CARBOHYD	751	751	N-LINKED (GLCNAC. . ) (POTENTIAL).	RN
PT CARBOHYD	754	754	N-LINKED (GLCNAC. . ) (POTENTIAL).	RN
PT CARBOHYD	900	900	N-LINKED (GLCNAC. . ) (POTENTIAL).	RN
SQ SEQUENCE	1429	AA;	157115 MW; 255BDYA62C025DB CRC64;	RN
Query Match	3.4%	Score 166.5; DB 1; Length 1429;		RX MEDLINE=21094508; PubMed=11182080;
Best Local Similarity	20.4%; Pred. No. 0.024;	Mismatches 113; Conservative 43; Gaps 22;		RN Tanigaki K., Nogaki F., Takahashi J., Tashiro K., Kurooka H.,
Or	438 CABCPGSWIKGYCZACNNASACDDGGSGNS--GSSRVIAGGGTSIGVHPWQ 494			RN Horjo T.,
Db	643 CSERA-----NDGNCDACTTACKDGGCGSKRPFKRY-----GNMCADF--			RN "Notch1 and Notch3 instructively restrict bfgf-responsive multipotent neural progenitor cells to an astroglial fate";
Qy	495 FGGGINSVSYCNQGCAN-----SMLADKICDQAQNVLSC 528			RN Neuron 29:45-55(2001).
Db	688 FANGV---CNGACNNECCLYDGMDLPAVRCPYKIRKHCASRFANGCDPENITNGC 742			RN [4]
Qy	529 GFDAGDCGQDFHFLYKVILLPNOTHYIIPKGCLPLVFSFAVKRGVEGAVSDNPIRH 588			RN TISSUE SPECIFICITY.
Db	743 GFDGGDC-----DNBT-----			RN MEDLINE=93202015; PubMed=1295745;
Qy	589 ASIANKWKTHLIMHSIGNATT-HFNITFQNTDEEK-----MQITWEVD 634			RN Weimaster G., Roberts V.J., Lemke G.;
Db	754 -----NATITNTRIVQ-MDPKEFOVTGQSISLMEISSALRVTRIQ 794			RN "Notch2: a second mammalian Notch gene.";
Qy	635 -TREGPKLNSTAQKGYENLVSPII---LPEAEILFEDIPKPKRFPKFKRDVNTRRA 689			RN Development 116:931-941(1992).
Db	795 RDEEGPLV--EONGESEMDRYKONERQLETEGHVLPSPISR----KIKRATNTGTVV 846			RN [5]
Qy	690 QEEVKPLVNISLIPKAQLSINTLDOIHEGDITLKQYNL-----KSALLRSL- 740			
Db	847 YLEVQENCDTGKCYLKAOSVUDSISAKLAKGKIDSIGIPISEALVAEPRKSGNTGFLS 906			
Qy	907 WNALLLIGAGCLIVVNLMLGALPGRNTRKRRMINASVMMPPMNEENKRKRHQSTISSQ 966			
Qy	759 NDSSIVAPQKQV-----HKSTLPSLNLGVSE-----RLQRLTPAVSKVNG 799			
Db	967 HSLLBASPVYKQRNELQHYSLYPNPOQYNGNDLGFDFNHTNLQDPTEPBPESPRL 1026			
Qy	800 HDQGONPPADLESTARFRVE---THQKT---IGGNVTKERPPSLIVPLESOMTKKKI 852			
Db	1027 HTEAGGSVYAITERPTRESVNTIDPRHRNTRVHLWIASNSAEEKSDLV--HEAKCIA 1082			
Qy	853 TGEKRNENRMEEN 865			
Db	1083 AGADVNMAMCDEN 1095			
RESULT 5				
ID NTCL_RAT				
NTCL_RAT STANDARD; PRT; 2531 AA.				
AC Q07008;				
DT 01-NOV-1995 (Rel. 32, Created)				
DT 15-JUL-1999 (Rel. 38, Last sequence update)				
DT 28-FEB-2003 (Rel. 41, Last annotation update)				
DE Neurogenic locus notch homolog protein 1 precursor (Notch 1).				
GN NORCH1.				
OS Rattus norvegicus (Rat).				
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
OC NCBI_TaxID=10116;				
RN [1] SEQUENCE FROM N.A.				
RP TISSUE=Schwann cell;				
RX MEDLINE=92111383; PubMed=1764995;				
RA Weimaster G., Roberts V.J., Lemke G.;				
RT A homolog of Drosophila Notch expressed during mammalian development.";				
RT Development 113:199-205(1991).				
RT [2]				
RT REVISIONS TO 1652-1653.				
RA Weimaster G.;				
RA Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.				
RA [3]				
RP FUNCTION.				
RN MEDLINE=21094508; PubMed=11182080;				
RN Tanigaki K., Nogaki F., Takahashi J., Tashiro K., Kurooka H.,				
RN Horjo T.,				
RT "Notch1 and Notch3 instructively restrict bfgf-responsive multipotent neural progenitor cells to an astroglial fate";				
RT Neuron 29:45-55(2001).				
RN [4]				
RP TISSUE SPECIFICITY.				
RN MEDLINE=93202015; PubMed=1295745;				
RN Weimaster G., Roberts V.J., Lemke G.;				
RN "Notch2: a second mammalian Notch gene.";				
RN Development 116:931-941(1992).				
RN [5]				

			DR	SMART; SMO0248; ANK; 6.
RK	MEDLINE=2131789; PubMed=11438922;		DR	SMART; SMO0179; EGF-CA; 25.
RA	Irvin D.K., Zurich S.D., Nguyen T., Weinmaster G., Kornblum H.I.; "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple functional roles for the Notch-DSL signaling system during brain development".		DR	SMART; SMO0009; NL; 2.
RT	J. Comp. Neurol. 436:167-181 (2001).		DR	PROSITE; PS0297; ANK REP REGION; 1.
RU	-!- FUNCTION: Functions as a receptor for membrane-bound ligands Jagged1, Jagged2 and Delta to regulate cell-fate determination.		DR	PROSITE; PS00088; ANK REPEAT; 4.
RC	Upon ligand activation through the released notch intracellular domain (NICD) it forms a transcriptional activator complex with RBP-J kappa and activates genes of the enhancer of split locus.		DR	PROSITE; PS00010; ASX-HYDROXYL; 22.
CC	Affects the implementation of differentiation, proliferation and apoptotic programs (BY similarity). Acts instructively to control the cell fate determination of CNS multipotent progenitor cells, resulting in astrogial induction and neuron/oligodendrocyte suppression.		DR	PROSITE; PS00027; EGF_1'; 35.
CC	-!- SUBUNIT: Heterodimer of a C-terminal fragment N(CT) and a N-terminal fragment N(EC), which are probably linked by disulfide bonds (BY similarity).		DR	PROSITE; PS01186; EGF_2'; 26.
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein. Following proteolytic processing NICD is translocated to the nucleus (BY similarity).		DR	PROSITE; PS0026; EGF_3'; 36.
CC	-!- TISSUE SPECIFICITY: Expressed in the brain, kidney and spleen.		KW	Developmental protein; Repeat; ANK repeat; Phosphorylation.
CC	Expressed in postnatal central nervous system (CNS) germinal zones and, in early postnatal life, within numerous cells throughout the CNS. Found in both subventricular and ventricular germinal zones.		KW	Transmembrane; Glycoprotein; Signal; SIGNAL; 18
CC	-!- DEVELOPMENTAL STAGE: In the embryo, highest levels occur between days 12 and 14 and decrease rapidly to much lower levels in the adult.		KW	NOTCH EXTRACELLULAR TRUNCATION (BY SIMILARITY).
CC	-!- PTM: Synthesized in the endoplasmic reticulum as an inactive form which is proteolytically cleaved by a furin-like convertase in the trans-Golgi network before it reaches the Plasma membrane to yield an active, ligand-accessible form. Cleavage results in a C-terminal fragment N(TM) and a N-terminal fragment N(EC). Following ligand binding, it is cleaved by TNF-alpha converting enzyme (TACE) to yield a membrane-associated intermediate fragment called notch extracellular truncation (NEXT). This fragment is then cleaved by presenilin dependent gamma-secretase to release a notch-derived peptide containing the intracellular domain (NICD) from the membrane (BY similarity).		KW	NOTCH INTRACELLULAR DOMAIN (BY SIMILARITY).
CC	-!- SIMILARITY: Belongs to the NOTCH family.		FT	EXTRACELLULAR (POTENTIAL).
CC	-!- SIMILARITY: Contains 3 Lin-Notch repeats.		FT	POTENTIAL.
CC	-!- SIMILARITY: Contains 5 ANK repeats.		FT	CYTOSPLASMIC (POTENTIAL).
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@isb-sib.ch).		FT	EGF-LIKE 1.
CC	-----		FT	EGF-LIKE 2.
CC	-----		FT	EGF-LIKE 3.
CC	-----		FT	EGF-LIKE 4.
CC	-----		FT	EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
CC	-----		FT	EGF-LIKE 6.
CC	-----		FT	EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
CC	-----		FT	EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
CC	-----		FT	EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
CC	-----		FT	EGF-LIKE 10.
CC	-----		FT	EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
CC	-----		FT	EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
CC	-----		FT	EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
CC	-----		FT	EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
CC	-----		FT	EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
CC	-----		FT	EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
CC	-----		FT	EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
CC	-----		FT	EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
CC	-----		FT	EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
CC	-----		FT	EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
CC	-----		FT	EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
CC	-----		FT	EGF-LIKE 22, CALCIUM-BINDING (POTENTIAL).
CC	-----		FT	EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
CC	-----		FT	EGF-LIKE 24.
CC	-----		FT	EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
CC	-----		FT	EGF-LIKE 26.
CC	-----		FT	EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
CC	-----		FT	EGF-LIKE 28.
CC	-----		FT	EGF-LIKE 29.
CC	-----		FT	CALCIUM-BINDING (POTENTIAL).
CC	-----		FT	EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
CC	-----		FT	EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
CC	-----		FT	EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
CC	-----		FT	EGF-LIKE 33, CALCIUM-BINDING (POTENTIAL).
CC	-----		FT	EGF-LIKE 34.
CC	-----		FT	EGF-LIKE 35.
CC	-----		FT	EGF-LIKE 36.
CC	-----		FT	LIN-NOTCH 1.
CC	-----		FT	LIN-NOTCH 2.
CC	-----		FT	LIN-NOTCH 3.
CC	-----		FT	ANK 1.
CC	-----		FT	ANK 2.
CC	-----		FT	ANK 3.
CC	-----		FT	ANK 4.
CC	-----		FT	ANK 5.
CC	-----		FT	POLY-ALA.
CC	-----		FT	POLY-GLU.
CC	-----		FT	POLY-PRO.
CC	-----		FT	POLY-SER.
CC	-----		FT	CLEAVAGE BY (FURIN-LIKE PROTEASE) (BY SIMILARITY).
CC	-----		FT	DISULFID.
PRINTS; PRO0011; EGFLMINTN.		24	37	

FT	DISULFID	31	46	BY SIMILARITY.
FT	DISULFID	48	57	BY SIMILARITY.
FT	DISULFID	63	74	BY SIMILARITY.
FT	DISULFID	68	87	BY SIMILARITY.
FT	DISULFID	89	98	BY SIMILARITY.
FT	DISULFID	106	117	BY SIMILARITY.
FT	DISULFID	111	127	BY SIMILARITY.
FT	DISULFID	129	138	BY SIMILARITY.
FT	DISULFID	144	155	BY SIMILARITY.
FT	DISULFID	149	164	BY SIMILARITY.
FT	DISULFID	166	175	BY SIMILARITY.
FT	DISULFID	182	195	BY SIMILARITY.
FT	DISULFID	189	204	BY SIMILARITY.
FT	DISULFID	215	215	BY SIMILARITY.
FT	DISULFID	222	233	BY SIMILARITY.
FT	DISULFID	227	243	BY SIMILARITY.
				Query Match 3.4%; Score 165; DB 1; Length 2531;
	Best Local Similarity 20.6%; Pred. No. 0 072; Mismatches 179; Indels 186; Gaps 27;			Matches 111; Conservative 63; MisMatches 179; Indels 186; Gaps 27;
OY	412 GKDWPDPSPYSHSKGQKVTLMWVPCAGCGCSWVKGYCDKACNSACWDGPDGSN	471	QGCA 510	SEQUENCE OF 731-1899 FROM N.A. (ISOFORM 1), AND DEVELOPMENTAL STAGE.
Db	1437 GRDIP-----POEAECLLCPQEDAGNK-----VCLQCNHACGDGGCSLN	1482	RA	RA Copeland N.G.; Gridley T.;
QY	472 SGSS-----RTYAGG----GGTGSIGVGPWQFGGGINSVSYCM----QGCA	510	RC	RC "Cloning, analysis, and chromosomal localization of Notch-1, a mouse homolog of Drosophila Notch.";
Db	1483 FNDPKWKNCTOSLOCWVYFSDGHGDSOCNSAGCLFDGPDQPLFEG----QNPPLYDQCK	1537	RT	RT Genomics 15:259-264(1993).
QY	511 NSVLADKCDQAGCNVLSGCFDAGDCGQDFHFE----LYKVILLP-----NQTHWYIP	558	RN	RN SEQUENCE OF 731-1899 FROM N.A. (ISOFORM 1), AND DEVELOPMENTAL STAGE.
Db	1538 DHF-SGHDQGNSAECWMDGDAEA-EHVPLAEGVGRGASDPTIHLASIAWKWTHLI	601	RA	RA MEDLINE=9305001; PubMed=14265644;
QY	559 KGBCI-----PYFSFALEVAKRGVGRGASDPTIHLASIAWKWTHLI 601	1595	RA	RA Reaume A.G., Conlon R.A., Zirngibl R., Yamaguchi T.P., Rossant J.;
Db	1596 VSHVLTAVWVKRDAQGOOMIFFY-----GREBELRKPIKEPSAV--GWAATSL	1644	RC	RC "Expression analysis of a Notch homologue in the mouse embryo.";
QY	602 -----MHSGMNTATHFLNTFQNTNDEEFKMQITWDTRE-----GP	639	RT	RT Dev. Biol. 159:377-387(1992).
Db	1645 PGTINGGRQRELDPMHDIGSIVI-----LEIDNRQCVOSSSQFOSATDVAA	1691	RN	RN TISSUE=Embryo;
QY	640 KUNSTAQKQEVENVSPIILPEAELFEDIP-----KE	672	RA	RA MEDLINE=93048835; PubMed=1425352;
Db	1692 FGGLAUSLGSNLNTPKTIAV-KSETVEPPLSQLHMLVAAAFVLLFFVGCGVLLSLRK	1750	RA	RA Franco del Amo F., Smith D.E., Swiatek P.J., Gendron-Maguire M.,
QY	673 KR----FPK-FKRDVNSTRAQEEVKIPL-VNTSLLP-KDAGLSLNTLQDIEHG-	721	RC	RC Greenspan R.J., McMahon A.P., Gridley T.;
Db	1751 RRHQHQLWPEGKIVSEASKKERE----PIGEDSVLGKPKNAQDGAQDNQNEWGD	1805	RT	RT "Expression pattern of Moch, a mouse homolog of Drosophila Notch, suggests an important role in early postimplantation mouse development.";
QY	722 -DTLKGYNLSKSALLRFMLMSOHAKKKNQAIITDETDNSLVARPQKQVHSILPNSLG	780	RN	RN Development 115:737-744(1992).
Db	1806 EDETTKKFRPEEPVVLPLDDQDTRHWTQHDLADLRVSAWAPPPQ----- 1854		RA	RA TISSUE=Embryo;
QY	781 VSSLRLQRTFPAVSKWNGHDPGQNPDPDLETTARFVETHHQKTGGGNVTEKEPKPSL	839	RC	RC SEQUENCE OF 1559-1673 FROM N.A.
Db	1855 -----GEVDADCMDVNVRGPD-GFTPLMIASCSCGGELBT-----GNSEEEEDAPAVI	1900	RT	RT MEDLINE=9336499; PubMed=10337798;
	RESULT 6		RA	RA Lee J.S., Ishimoto A., Yangawa S.I.;
	NRC1_MOUSE		RT	RT "Murine leukemia provirus-mediated activation of the Notch1 gene leads to induction of Hes-1 in a mouse T lymphoma cell line, DL-3.";
ID	_NRC1_MOUSE		RT	RT FEBS Lett. 455:276-280(1999).
AC	Q06007; Q61905; Q9QJC2; Q9QWS8; Q9R0X7;		RN	RN [5]
DT	01-NOV-1995 (Rel. 32, Created)		RA	RA SEQUENCE OF 1559-1673 FROM N.A.
DT	01-FEB-1996 (Rel. 33, Last sequence update)		RA	RA MEDLINE=936499; PubMed=10337798;
DT	10-OCT-2003 (Rel. 42, Last annotation update)		RA	RA EXP. Cell Res. 204:364-372(1993).
DE	Neurogenetic locus notch homolog protein 1 precursor (Notch 1) (Notch A) (mT1) (p300)		RA	RA SEQUENCE OF 1559-1673 FROM N.A.
DE	NOTCH1 OR NOTCH.		RA	RA MEDLINE=98029996; PubMed=934671;
OS	Mus musculus (Mouse).		RA	RA Messerle W., Poller M., Nehls M., Eggert H., Boehm T.;
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		RT	RT "Dynamic changes in gene expression during in vitro differentiation of mouse embryonic stem cells.";
NCBItaxID=10900;	[1]		RA	RA Cytokines Cell. Mol. Ther. 1:139-143(1995).
RN			RA	RA [7]
			RA	RA SEQUENCE OF 1655-1659, CLEAVAGE BY FURIN-LIKE CONVERTASE, AND MUTAGENESIS OF 1651-ARG->Arg-1654.
			RA	RA MEDLINE=98318619; PubMed=9831487;
			RA	RA Logeat F., Bessia C., Brou C., LeBail O., Jarriault S., Seidah N.G., Israel A.;
			RA	RA "The Notch1 receptor is cleaved constitutively by a furin-like convertase.";
			RA	RA Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112(1998).
			RA	RA [8]
			RA	RA PARTIAL SEQUENCE, AND POST-TRANSLATIONAL PROCESSING.
			RA	RA MEDLINE=21523956; PubMed=11518718;
			RT	RT "Murine notch homolog (N1-4) undergo presenilin-dependent proteolysis.";
			RT	RT J. Biol. Chem. 276:40268-40273(2001).
			RA	RA [9]
			RA	RA POST-TRANSLATIONAL PROCESSING.
			RA	RA MEDLINE=2137376; PubMed=11459911;
			RA	RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
			RT	RT "Conservation of the biochemical mechanisms of signal transduction among mammalian Notch family members.";
			RT	RT Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
			RA	RA [10]
			RA	RA INTERACTION WITH DTX1 AND DTX2.
			RA	RA MEDLINE=21122790; PubMed=11226722;



QY	747 KIRKQAITTDETNUSLVAPQEKOVHKSIILPNSLGVSRLLQRUTPAVSKVNGHPOQNP 806	DR Pfam; PF00023; ankyrin repeat; 6.
Db	1832 QWTOQHHLRADRSAMAFPPQ-----GEVADMDMVNRGP-----GFT 1873	DR Pfam; PRO0008; EGF; 36.
QY	807 PDLLETTARPRVTHTKIGVNTKEKPPSII-----VPLESQTKEKIT----- 853	DR PIRSF02279; Notch; 3.
Db	1874 PLMIAASCGGGLT-----GNSEEEEDAPAVISDFIYQGASLNQNDRTGETALHAR 1927	DR PRINTS; PRO0010; EGFBLOOD.
QY	854 -GEKENSREMEAAENIT-----GVTEDLU 877	DR PRINTS; PRO0011; EGFLAMININ.
Db	1928 YSRSDRRIKLEASADANICDDNGRTPHAASDAGCQVFQIL 1970	DR PRINTS; PRO1452; NOTCH.
<b>RESULT 7</b>		
NTC1_BRAE	STANDARD;	PRT; 2437 AA.
ID	_NWC1_BRAE	PRT; PS50297; ANK REP REPEAT; 1.
AC	P46530;	DR PROSITE; PS50088; ANK REPEAT; 4.
DT	01-NOV-1995 (Rel. 32, Last sequence update)	DR PROSITE; PS0001; ASX_HYDROXYL; 23.
DT	28-FEB-2003 (Rel. 41, Last annotation update)	DR PROSITE; PS00022; EGF 1; 34.
DE	Neurogenic locus notch homolog protein 1 precursor.	DR PROSITE; PS00186; EGF 2; 28.
GN	NOTCH1A OR NOTCH	DR PROSITE; PS50026; EGF 3; 36.
OS	Brachydanio rerio (Zebrafish) (Danio rerio).	DR PROSITE; PS00187; EGF CA; 22.
OC	Brachydanio; Metazoa; Chordata; Vertebrata; Euteleostomi; Cetaceans; Cyprinidae; Danio.	DR PROSITE; PS50027; ANK REP REGION; 1.
OX	NCBI_TaxID=955;	DR PROSITE; PS50027; ANK REP REGION; 1.
RN	[1]	DR PROSITE; PS50027; ANK REP REGION; 1.
RP	SEQUENCE FROM N.A.	DR PROSITE; PS50027; ANK REP REGION; 1.
RC	TISSUE=Embryo;	DR PROSITE; PS50027; ANK REP REGION; 1.
RX	MEDLINE=94128602; PubMed=8277791;	DR PROSITE; PS50027; ANK REP REGION; 1.
RA	Bierkamp, C.; Campos-Ortega, J.A.;	DR PROSITE; PS50027; ANK REP REGION; 1.
RT	"A zebrafish homologue of the Drosophila neurogenic gene Notch and its pattern of transcription during early embryogenesis.";	DR PROSITE; PS50027; ANK REP REGION; 1.
RL	Mech. Dev. 43:87-100(1993).	DR PROSITE; PS50027; ANK REP REGION; 1.
CC	-I- FUNCTION: Implicated in cell fate specifications during embryo development. May be involved in the formation of the neural plate, notochord and brain vesicles.	DR PROSITE; PS50027; ANK REP REGION; 1.
CC	-I- SUBCELLULAR LOCATION: Type I membrane protein.	DR PROSITE; PS50027; ANK REP REGION; 1.
CC	-I- DEVELOPMENTAL STAGE: Expressed in all cells in pregastrulation stages. During gastrulation, expression is differentially expressed, accumulating predominantly in the presomitic mesoderm and notochord. At the end of gastrulation, expressed along the anterior-posterior axis including the developing neural plate and differentiating mesoderm. Also present in the developing brain and head regions.	DR PROSITE; PS50027; ANK REP REGION; 1.
CC	-I- SIMILARITY: Belongs to the NOTCH family.	DR PROSITE; PS50027; ANK REP REGION; 1.
CC	--I- SIMILARITY: Contains 36 EGF-like domains.	DR PROSITE; PS50027; ANK REP REGION; 1.
CC	--I- SIMILARITY: Contains 3 Lnn/Notch repeats.	DR PROSITE; PS50027; ANK REP REGION; 1.
CC	--I- SIMILARITY: Contains 6 ANK repeats.	DR PROSITE; PS50027; ANK REP REGION; 1.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@isb-sib.ch).	DR PROSITE; PS50027; ANK REP REGION; 1.
CC	EMBL; X69088; CKA4881.1; -.	DR Pfam; PF00023; ankyrin repeat; 6.
DR	PIR; S42612; S22612.	DR Pfam; PRO0008; EGF; 36.
DR	HSSP; P0740; 1EDW.	DR PIRSF02279; Notch; 3.
DR	ZFIN; ZDB-GENE:990415-173; notch1a.	DR PRINTS; PRO0010; EGFBLOOD.
DR	InterPro; IPR002110; ANK.	DR PRINTS; PRO0011; EGFLAMININ.
DR	InterPro; IPR000152; Asx_hydroxyl_S.	DR PRINTS; PRO1452; NOTCH.
DR	InterPro; IPR000742; BGF_2.	DR SMART; SW00248; ANK; 6.
DR	InterPro; IPR001881; BGF_Ca.	DR SMART; SW00179; EGF_CA; 22.
DR	InterPro; IPR001438; EGF_II.	DR SMART; SW00044; NLJ; 3.
DR	InterPro; IPR006209; BGF_Like.	DR SMART; SW00044; NLJ; 3.
DR	InterPro; IPR002049; Laminin_EGF.	DR SMART; SW00044; NLJ; 3.
DR	InterPro; IPR000297; Notch.	DR SMART; SW00044; NLJ; 3.
DR	InterPro; IPR000800; Notch_dom.	DR SMART; SW00044; NLJ; 3.
FT	SITE	DR SITE

Query	Matched 119; Conservative 58; Mismatches 201; Index 197; Gaps 29;
Best Local Similarity	3.3%; Score 162.5; DB 1; Length 2437;
Pred. No. 0.097;	
Db	421 YSHSKGQ-----KVLTWPPNCAGCPGSWIKUDYCKACNNASACDAGGDSGNS 472
Qy	1427     :    :    :    :    :    :    :    :    :    :    :    :    :
Db	1478 -----SLNFDPPWQNCSSALQCWRYFNDG----KCDPQCATAGLYCGFDC 1519
Qy	473 GGSRVIAGGGGTGSIIVGSHWPQ-----FGGINSVSYCNGCANS----- 512
Db	1520 ORLEGQCNPLDYQYCRDHYADGHCDGCGNAECWBDGICADDVPPQKLAVSSILVWHIP 1579
Qy	551 -----NQTHYI-----PKGCL--PFSSFAEVARGVEGAYSDHPIR 587
Db	1580 PDELNRSSSFLRLSSLIHTNVFRRDANGEALIFPYG-----SEHLISK 1626
Qy	568 HASIANKWTKTIHIMHSQMNATIHFN-LTFQNTNDDEPKMOT-WEDTDE----- 637
Db	1627 HK--RSWDTPGGLIMORARRSLTSFLKRTRRELDMEVKSVIVLBDNROCFQOSDEC 1684
Qy	369 -----GPKINSTAKQSYENLVSPITLPERELFDPKPKRKFPKRHDVNSTRAQ 690
Db	638 -----EMYMFVLIAL---AVLAALAVVUVSRKKRKGHQQLWFP-----EGKYNEP 1769
Qy	1685 FQSATDVAAFLGLASSG--NLNUVP-----YIAUTVSQQGPK-----TG 1723
Db	691 BEVKPLVANISLIPKAQDLSNLTD-----OLEHDITLGYNNSKSALLRSFJMSQ 744
Qy	1724 EMYMFVLIAL---AVLAALAVVUVSRKKRKGHQQLWFP-----KPRPLIVPESOMTKEKK 851
Db	745 HAKINQOAITDTNDSHLVAPOEKQVKMSTLNSL---GVSERLQRLTPAVSV-KVNG- 799
Db	1770 KCKGRKEPV-----GBDSVGLKPLKLNSSLSMBOLSEWAEDDTINKPRFEGOSILEMSGQ 1824
Qy	800 --HGCGQNPLD-ETTAFFRVELHTKTTGNYVTE-----KPRPLIVPESOMTKEKK 851
Db	1825 LDHRQWTQHLD---AADLRLNSMAPTPQGQLENCDMDVNVRGPDPGFTPL---MASCS 1878
Qy	852 ITGKEKENSMEENAENHIGVTEVULGRKLOHYD 886
Db	1879 GGSLNENEGEAEEDPSADVITDPYHGANLHNOD 1913
RESULT 8	
NTC3_HUMAN STANDARD; PRT; 2321 AA.	
ID	NTC3_HUMAN STANDARD; PRT; 2321 AA.
AC	Q9B447; Q9UEB3; Q9UPL3; Q9Y6L8;
DT	28-FEB-2003 (Rel. 41, Created)
DT	28-FEB-2003 (Rel. 41, Last sequence update)
DT	28-FEB-2003 (Rel. 41, Last annotation update)
DB	Neurogenic locus notch homolog protein 3 precursor (Notch 3).
GN	NOTCH3.
OS	Homo sapiens (Human).
OC	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
PT	SEQUENCE FROM N.A.
PT	MEDLINE=97032728; PubMed=8878478;
PT	Jouvel A., Corpechot C., Ducros A., Vahedi K., Chabriat H., Mouton P., Alarconich S., Domenga V., Cecillion M., Mareschal J., Vaysseire C., Cnaud C., Cabanis E.-A., Ruchoux M.M., Weissenbach J., Bach J.-P., Bousser M.-G., Tourrier-Lasserre E.;
PT	"Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke and dementia.,"
PT	Nature 383:707-710(1996).
PT	[2]
PT	SEQUENCE FROM N.A.
PT	Gurel M., Artavaslis-Tsakonas S., Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
PT	[3]
PT	SEQUENCE FROM N.A.
PT	Lauterdin J.E., McCready P.M., Skowronski E., Adamson A.W.,
RA	
RN	
RA	

RA	Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stilwagen S.,	CC	This SWISS-PROT entry is copyright
RA	Pham H., Velasco N., Barnes J., Dangaran I., Poundstone P., Andreise T.,	CC	between the Swiss Institute of Bioin
RA	Christensen M., Georgescu A., Avila J., Liu S., Attix C., Andreise T.,	CC	the European Bioinformatics Instituti
RA	Frankenheim M., Amico-Keller G., Coeffield J., Duarte S., Lucas S.,	CC	use by non-profit institutions as
RA	Bruce R., Thomas P., Guan C., Krommiller B., Arellano A., Olsen A.S.,	CC	modified and this statement is not
RA	Montgomery M., Ow D., Nolan M., Trong S., Kobayashi A., Olsen A.S.,	CC	entities requires a license agreement
RA	Carrano A.V.;	CC	or send an email to license@ibsl-sib.ch.
RT	"sequence analysis of an 1.5 Mb olfactory receptor (OLFR) cluster in	CC	
RT	1993.1.,"	CC	
RL	Submitted (May-1998) to the EMBL/Genbank/DBJ databases.	CC	
RN	[4] VARIANTS CADASIL TYR-49; CYS-71; CYS-90; CYS-110; CYS-133; CYS-141;	DR	EMBL: U97669; AAB91371.1; -
RP	ARG-146; CYS-153; CYS-169; CYS-171; CYS-182; ARG-185; SSR-212;	DR	EMBL: AF058900; AAC13461.1; -
RP	GLY-214; TRP-224; CYS-248; TYR-542; CYS-558; CYS-578; CYS-728;	DR	EMBL: AF058883; AAC13461.1; JOINED.
RP	CYS-985; CYS-1006; CYS-1031; CYS-1231 AND ARG-1261, AND VARIANTS	DR	EMBL: AF058883; AAC13461.1; JOINED.
RP	ARG-170; LEU-179; GLN-1133; MET-1183 AND ALA-2223.	DR	EMBL: AF058885; AAC13461.1; JOINED.
RX	MEDLINE=9809753; PubMed=988399;	DR	EMBL: AF058886; AAC13461.1; JOINED.
RA	Joutel A., Vahedi K., Corpauchot C., Troesch A., Chabriat H.,	DR	EMBL: AF058887; AAC13461.1; JOINED.
RA	Vaysiere C., Crouaud C., Maciazeck J., Weissenbach J., Bouller M.-G.,	DR	EMBL: AF058888; AAC13461.1; JOINED.
RA	Bach J.-F., Tournier-Lasserre E.,	DR	EMBL: AF05889; AAC13461.1; JOINED.
RT	"Strong clustering and stereotyped nature of Notch3 mutations in	DR	EMBL: AF058890; AAC13461.1; JOINED.
RT	CADASIL patients;"	DR	EMBL: AF058891; AAC13461.1; JOINED.
RL	Lancet 350:1511-1515(1997).	DR	EMBL: AF058892; AAC13461.1; JOINED.
RN	[5] VARIANT CADASIL 114-GLY--PRO-120 DEL.	DR	EMBL: AF058893; AAC13461.1; JOINED.
RX	MEDLINE=2056473; PubMed=0802807;	DR	EMBL: AF058894; AAC13461.1; JOINED.
RA	Joutel A., Chabriat H., Domenga V., Vaysiere C., Tournier-Lasserre E.,	DR	EMBL: AF058895; AAC13461.1; JOINED.
RA	Ruchoux M.M., Lucas C., Leyv D., Bousser M.-G., Tournier-Lasserre E.,	DR	EMBL: AF058896; AAC13461.1; JOINED.
RT	"Splice site mutation causing a seven amino acid Notch3 in-frame	DR	EMBL: AF058897; AAC13461.1; JOINED.
RT	deletion in CADASIL."	DR	EMBL: AF058898; AAC13461.1; JOINED.
RL	Neurology 54:1874-1875(2000).	DR	EMBL: AF058899; AAC13461.1; JOINED.
RN	[6] IDENTIFICATION OF LIGANDS.	DR	EMBL: AC004257; AAC04663; AAC15789.1; ALT_INITI
RX	MEDLINE=99180765; PubMed=10079256;	DR	EMBL: AC004257; AAC04897.1; -
RA	Gray G.B., Mann R.S., Mitsiadis E., Henrique D., Carcangioli M.-L.,	DR	HSSP; P00740; IEDM.
RA	Banks A., Leiman J., Ward D., Ish-Horowicz D., Artavanis-Tsakonas S.,	DR	PIR; S7859; S78549.
RA	"Human ligands of the Notch receptor.";	DR	PIR; S7859; S78549.
RL	Am. J. Pathol. 154:783-794(1999).	DR	P00740; IEDM.
CC	- - FUNCTION: Functions as a receptor for membrane-bound ligands	DR	HSSP; P00740; IEDM.
CC	Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.	DR	Genew; HGNC:7083; NOTCH3.
CC	Upon ligand activation through the released notch intracellular	DR	NM: 125310; -
CC	domain (NICD) it forms a transcriptional activator complex with	DR	InterPro; IPR002110; ANK.
CC	RBP-J kappa and activates genes of the enhancer of split locus.	DR	InterPro; IPR00152; ASK_hydroxyl_S.
CC	Affects the implementation of differentiation, proliferation and	DR	InterPro; IPR00742; EGF_2.
CC	apoptotic programs (By similarity).	DR	InterPro; IPR01881; EGF_Ca.
CC	- - SUBUNIT: Heterodimer of a C-terminal fragment (NICD) and a N-	DR	InterPro; IPR01438; EGF_II.
CC	terminal fragment N(TM) and a N-terminal fragment N(EC) which are probably linked by disulfide	DR	InterPro; IPR002049; EGF_Like.
CC	bonds (By similarity).	DR	InterPro; IPR008297; Notch.
CC	- - SUBCELLULAR LOCATION: Type I membrane protein. Following	DR	InterPro; IPR000800; Notch_dom.
CC	proteolytic processing NICD is translocated to the nucleus.	DR	PFAM; PF00008; EGF_3.
CC	- - TISSUE SPECIFICITY: Ubiquitously expressed in fetal and adult	DR	PFAM; PF00066; notch_3.
CC	tissues.	DR	PFAM; PF000279; Notch_1.
CC	- - PTM: Synthesized in the endoplasmic reticulum as an inactive form	DR	PRINTS; PRO0010; EGFBLOOD.
CC	which is proteolytically cleaved by a furin-like convertase in the	DR	PRINTS; PRO0011; EGFLAMIN.
CC	trans-Golgi network before it reaches the plasma membrane to yield	DR	PRINTS; PRO01422; NOTCH.
CC	an active, ligand-accessible form. Cleavage results in a C-	DR	SMART; SM00248; ANK_6.
CC	terminal fragment N(TM) and a N-terminal fragment N(EC). Following	DR	SMART; SM00179; EGF_Ca_19.
CC	ligand binding, it is cleaved by TNF-alpha converting enzyme	DR	SMART; SM00044; NL_3.
CC	(TACE) to yield a membrane-associated intermediate fragment called	DR	PROSITE; PS50297; ANK_REPEAT_1.
CC	notch extracellular truncation (NEXT). This fragment is then	DR	PROSITE; PS50088; ANK_RPEAT_4.
CC	cleaved by present in dependent gamma-secretase to release a	DR	PROSITE; PS50010; ASX_HYDROXYL_18.
CC	notch-derived peptide containing the intracellular domain (NICD)	DR	PROSITE; PS50022; EGF_1_33.
CC	from the membrane (By similarity).	DR	PROSITE; PS50186; EGF_2_25.
CC	- - PTM: Phosphorylated (By similarity).	DR	PROSITE; PS50026; EGF_3_34.
CC	- - DISEASE: Defects in NOTCH3 are associated with cerebral autosomal	DR	PROSITE; PS50187; EGF_CA_16.
CC	dominant arteriopathy with subcortical infarcts and	DR	PROSITE; PS50188; NOTCH_E.
CC	leukocephalopathy (CADASIL)	KW	Developmental protein; Repeat; ANK_r
CC	of stroke and dementia of which key features include recurrent	KW	Transmembrane; Glycoprotein; Signal;
CC	subcortical ischemic events and vascular dementia.	KW	Disease mutation.
CC	- - SIMILARITY: Belongs to the NOTCH family.	FT	SIGNAL_1_39; POTENTI
CC	- - SIMILARITY: Contains 34 EGF-like domains.	FT	NEURONE
CC	- - SIMILARITY: Contains 3 Lnn/Notch repeats.	FT	NOTCH_E
CC	- - SIMILARITY: Contains 5 ANK repeats.	FT	SIMILAR
CC		FT	NOTCH_I
CC		CHAIN	SIMILAR
CC		CHAIN	1629
CC		CHAIN	2321
CC		CHAIN	1662
CC		CHAIN	2321

RP	SEQUENCE FROM N.A.
RC	TISSUE=Lung, and Testis;
RX	MEDLINE=96281668; PubMed=8681805;
RA	Uytterhaeve H., Marzzi G., Wu G., Yan Q., Sascon D., Kitajewski J.,
RT	"Notch4/int-3, a mammalian proto-oncogene, is an endothelial cell-specific mammalian Notch gene";
RT	Development 122:2251-2259(1996).
RL	[4]
RN	SEQUENCE FROM N.A.
RP	Rowen L., Mahairas G., Qin S., Ahearn M.E., Dankers C., Lasky S.L.,
RA	Ioretz C., Schmidt S., Tipton R., Traiforoff R., Zackrone K., Hood L.;
RT	"Sequence of the mouse major histocompatibility locus Class III region.>"; Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
RL	[5]
RN	SEQUENCE OF 1436-1600 FROM N.A.
RP	MEDLINE=9925212; PubMed=10233982;
RA	Ide J.-S., Haruna T., Ishimoto A., Honjo T., Yanagawa S.-I.;
RT	"Intracellular type A particle-mediated activation of the Notch4/int3 gene in a mouse mammary tumor: generation of truncated Notch4/int3 mRNAs by retroviral splicing events.";
RT	J. Virol. 73:5166-5171(1999).
RN	[6]
RP	FUNCTION.
RX	MEDLINE=212244637; PubMed=11343305;
RA	Duttenhofer H., Ho J., Rossant J., Kitajewski J.;
RT	"Muscular patterning defects associated with expression of activated Notch4 in embryonic endochelium.";
RL	[7]
RN	Proc. Natl. Acad. Sci. U.S.A. 98:5643-5648 (2001).
RP	SEQUENCE OF 1463-1964, POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF VAL-1433.
RX	MEDLINE=21523956; PubMed=1158718;
RA	Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
RT	"Murine notch homologs (N1-4) undergo presenilin-dependent proteolysis.";
RL	J. Biol. Chem. 276:40268-40273 (2001).
RN	[8]
RP	POST-TRANSLATIONAL PROCESSING.
RX	MEDLINE=2137436; PubMed=11459941;
RA	Mizutani T., Taniguchi Y., Asai T., Hashimoto N., Honjo T.;
RT	"Conservation of the biochemical mechanisms of signal transduction among mammalian Notch family members.";
RL	Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
CC	-!- FUNCTION: Functions as a receptor for membrane-bound ligands Jagged1, Jagged2 and Delta to regulate cell-fate determination.
CC	Upon ligand activation through the released notch intracellular domain (NICD) it forms a transcriptional activator complex with RBP-J kappa and activates genes of the enhancer of split locus.
CC	Affects the implementation of differentiation, proliferation and apoptotic programs (By similarity). May regulate branching morphogenesis in the developing vascular system.
CC	-!- SUBUNIT: heterodimer of a C-terminal fragment N( <sup>rm</sup> ) and a N-terminal fragment NICD which are probably linked by disulfide bonds.
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein. Following proteolytic processing NICD is translocated to the nucleus.
CC	-!- TISSUE SPECIFICITY: Highly expressed in lung, moderately in heart, kidney, and at lower levels in the ovary and skeletal muscle. A very low expression is seen in the brain, intestine, liver and testis.
CC	-!- DEVELOPMENTAL STAGE: Highly expressed in endothelial cells during embryonic development from 9.0 dp.
CC	-!- PM: Synthesized in the endoplasmic reticulum as an inactive form which is proteolytically cleaved by a furin-like convertase in the trans-Golgi network before it reaches the plasma membrane to yield an active, ligand-accessible form. Cleavage results in a C-terminal fragment N( <sup>rm</sup> ) and a N-terminal fragment N(EC). Following ligand binding it is cleaved by tNF-alpha converting enzyme (TACE) to yield a membrane-associated intermediate fragment called notch extracellular truncation (NEXT). This fragment is then cleaved by presenilin dependent gamma-secretase to release a notch-derived peptide containing the intracellular domain (NICD).
RN	[1]
RP	REVISIONS, SEQUENCE FROM N.A.
RX	MEDLINE=97244599; PubMed=9150355;
RA	Robbins J., Blondel B.J., Gallahan D., Callahan R.;
RT	"Mouse mammary tumor gene int-3: a member of the notch gene family transforms mammary epithelial cells.";
RL	J. Virol. 66:2594-2599(1992).
RN	[2]
RP	REVISIONS, SEQUENCE FROM N.A.
RX	MEDLINE=92194507; PubMed=1312643;
RA	Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;	
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=92194507; PubMed=1312643;
RA	"Mouse mammary tumor gene int-3: a member of the notch gene family transforms mammary epithelial cells.";
RT	"The mouse mammary tumor gene INT3 is a unique member of the NOTCH gene family (NOTCH4)." ;
RL	Oncogene 14:1883-1890(1997).
RN	[3]

CC from the membrane.  
 CC  
 CC -!- PTM: Phosphorylated.  
 CC -!- DISASE: Loss of the extracellular domain causes constitutive activation of the Notch protein, which leads to hyperproliferation of glandular epithelial tissues and development of mammary carcinomas.  
 CC -!- SIMILARITY: Belongs to the NOTCH family.  
 CC -!- SIMILARITY: Contains 29 EGF-like domains.  
 CC -!- SIMILARITY: Contains 3 Lin/Notch repeats.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).  
 CC  
 CC  
 DR EMBL; M80456; AAB38377.1; -.  
 DR EMBL; U43691; AAC52631.1; -.  
 DR EMBL; AF030001; AAB8204.1; -.  
 DR EMBL; AB016772; BAA32281.1; ALT\_INIT.  
 DR EMBL; AB016773; BAA32284.1; ALT\_INIT.  
 DR EMBL; AB016774; BAA32285.1; -.  
 DR PIR; A38072; TWRVTF.  
 DR PIR; T09059; T09059.  
 DR HSSP; P08709; IBF9.  
 DR MGD; MGII:07471; Notch4.  
 DR Interpro; IPR002110; ANK.  
 DR Interpro; IPR00152; ASX\_Hydroxy1\_S.  
 DR Interpro; IPR000742; EGF\_2.  
 DR Interpro; IPR001881; EGF\_Ca.  
 DR Interpro; IPR01438; EGF\_T1.  
 DR Interpro; IPR002029; EGF\_Like.  
 DR Interpro; IPR002049; Laminin\_EGF.  
 DR Interpro; IPR008297; Notch.  
 DR Interpro; IPR000800; Notch\_dom.  
 DR Pfam; PF00023; ankyrin repeat; 6.  
 DR Pfam; PF00008; EGF; 27.  
 DR Pfam; PF00066; notch; 2.  
 DR Pfam; PF000279; Notch; 1.  
 DR Prints; PRO0010; EGFBLOOD.  
 DR Prints; PRO0011; EGFLAMININ.  
 DR Prints; PRO1452; NOTCH.  
 DR SMART; SMC0248; ANK; 6.  
 DR SMART; SMC0177; EGF\_Ca; 11.  
 DR SMART; SMC0044; NLN; 2.  
 PROSITE; PS50297; ANK REP REGION; 1.  
 DR PROSITE; PS50086; ANK\_REPEAT; 5.  
 PROSITE; PS00010; ASYHYDROXYL; 11.  
 DR PROSITE; PS00022; EGF; 1; 28.  
 PROSITE; PS01185; EGF; 2; 21.  
 PROSITE; PS50026; EGF; 3; 27.  
 DR PROSITE; PS01187; EGF\_Ca; 9.  
 DR Receptor; Transmembrane regulation; Activator; Differentiation; Developmental Protein; Repeat; EGF-like domain; Transmembrane; Glycoprotein; Signal; Phosphorylation; Proto-oncogene.  
 FT SIGNAL 1 20  
 FT CHAIN 21 1964 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 4.  
 FT CHAIN 1411 1964 TRANSFORMING PROTEIN INT-3.  
 FT CHAIN 1428 1964 NOTCH EXTRACELLULAR TRUNCATION.  
 FT CHAIN 1463 1964 NOTCH INTRACELLULAR DOMAIN.  
 FT DOMAIN 21 1943 TRANSMEM 1444 1964 POTENTIAL.  
 FT DOMAIN 1465 1964 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 61 1960 EGF-LIKE 1.  
 FT DOMAIN 115 1952 EGF-LIKE 3.  
 FT DOMAIN 153 1989 EGF-LIKE 4.  
 FT DOMAIN 191 229 EGF-LIKE 5; CALCIUM-BINDING (POTENTIAL).

FT	DOMAIN	231	271	EGF-LIKE 6.
FT	DOMAIN	273	309	EGF-LIKE 7; CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	311	350	EGF-LIKE 8; CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	352	388	EGF-LIKE 9; CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	389	427	EGF-LIKE 10.
FT	DOMAIN	429	470	EGF-LIKE 11; CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	472	508	EGF-LIKE 12; CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	510	546	EGF-LIKE 13; CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	548	584	EGF-LIKE 14; CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	586	622	EGF-LIKE 15; CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	623	656	EGF-LIKE 16.
FT	DOMAIN	658	686	EGF-LIKE 17.
FT	DOMAIN	688	724	EGF-LIKE 18.
FT	DOMAIN	726	762	EGF-LIKE 19.
FT	DOMAIN	764	800	EGF-LIKE 20.
FT	DOMAIN	803	839	EGF-LIKE 21.
FT	DOMAIN	841	877	EGF-LIKE 22.
FT	DOMAIN	878	924	EGF-LIKE 23.
FT	DOMAIN	926	962	EGF-LIKE 24.
FT	DOMAIN	964	1000	EGF-LIKE 25.
FT	DOMAIN	1002	1040	EGF-LIKE 26.
FT	DOMAIN	1042	1081	EGF-LIKE 27.
FT	DOMAIN	1083	1122	EGF-LIKE 28.
FT	DOMAIN	1126	1167	EGF-LIKE 29.
FT	REPEAT	1168	1208	LIN/NOTCH 1.
FT	REPEAT	1209	1242	LIN/NOTCH 2.
FT	REPEAT	1243	1282	LIN/NOTCH 3.

Query Match Similarity 28.1%; Score 152; DB 1; Length 1964;  
 Matches 41; Conservative 11; Pred. No. 0; 3 AA. Mismatches 32; Indels 62; Gaps 8;

QY	434	PVPNC---AEGERPGSNWIKRDGKDACKNNASCDWDGGDCSGNSGGSYIAGGGTSIGV 489
Db	1162	PGFRCQRPGASGEGRGR-GDGTODAGSGPGGDWDGADC-----SLGV 1203
QY	490	GHWQFGEGGINNTSYGC-----NGCA-----NSWLAQKF- 518
Db	1204	PDPPR--GCCPHSQCWILFRDRGHCQDCSBCLFQGYDCIPLTCIPAVDQYCRHFH 1260
QY	519	--CDQAQCNVLSCGFDGDC--GQD 538
Db	1261	NGHEEKGCINNAECGWDG3DCRPEGED 1286

RESULT 10

ID	NTC4_HUMAN	STANDARD,	PRT;	2003 AA.
AC	Q99466; 000306; Q99458; Q99940; Q9HJS8; Q9UIJ0;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DE	Neurogenic locus notch homolog protein 4 precursor (Notch 4) (Notch4).			
GN	NOTCH4.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.			
OX	NCBI_TAXID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A. (ISOFORM 1), AND POLYMORPHISM OF POLY-LEU.			
RC	SEQUENCE FROM N.A. (ISOFORM 1), AND POLYMORPHISM OF POLY-LEU.			
RX	TISSUE-Placental; MEDLINE=97311416; Pubmed=168133;			
RA	Sugaya K., Sasakura J., Nohata J., Kimura T., Fukagawa T., Nakamura Y., Ando A., Inoko H., Ikeura T., Mita K.; "Gene organization of human NOTCH4 and (CTG)n polymorphism in this gene [189:235-244 (1997)]."			
RL	human counterpart gene of mouse proto-oncogene Int3.";			
RN	[2]			
RP	SEQUENCE FROM N.A. (ISOFORMS 1, 2 AND 3).			
RC	TISSUE-Bone marrow, and Heart; MEDLINE=98360091; Pubmed=693032;			
RX	Li L., Huang G.M., Banta A.B., Deng Y., Smith T., Dong P., RA			

Friedman C., Chen L., Trask B.J., Spies T., Rowen L., Hood L.;  
 RT "Cloning, characterization, and the complete 56.8-kilobase DNA  
 sequence of the human NOTCH4 gene.",  
 RTR genomics 51:45-58(1998).  
 RN [3]  
 RP SEQUENCE OF 1-503 FROM N.A., AND VARIANTS GLN-117 AND GLN-317.  
 RA Miyagawa T., Tokunaga K., Hojo H.;  
 RT "Human notch4 gene variant";  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.  
 RN [4]  
 RP IDENTIFICATION OF LIGANDS.  
 MMLINE=99180765; PubMed=0079256;  
 RA Gray G.E., Mann R.S., Mitsidis E., Henrique D., Carcangiu M.-L.,  
 Banks A., Leiman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;  
 RT "Human ligands of the Notch receptor";  
 RL Am. J. Pathol. 154:785-794(1999).  
 CC [5]  
 CC FUNCTION: Functions as a receptor for membrane-bound ligands.  
 CC Jagged1, Jagged2 and Delta1 to regulate branching morphogenesis in the  
 developing vascular system (By similarity!).  
 CC Upon ligand activation through the released notch intracellular  
 domain (NICD) it forms a transcriptional activator complex with  
 RBP-J kappa and activates genes or the enhancer of split locus.  
 CC Affects the implementation of differentiation, proliferation and  
 apoptotic programs. May regulate branching morphogenesis in the  
 developing vascular system (By similarity!).  
 CC SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-  
 terminal fragment N(EC) which are probably linked by disulfide  
 bonds (By similarity). Type I membrane protein. Following  
 proteolytical processing NICD is translocated to the nucleus.  
 -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=3;  
 CC Isoform=Experimental confirmation may be lacking for some  
 CC Isoform;  
 CC IsoId=Q99466-1; Sequence=Displayed;  
 CC IsoId=099466-2; Sequence=VSP\_001406;  
 CC Name=3;  
 CC IsoId=099466-3; Sequence=VSP\_001407;  
 CC -!- TISSUE SPECIFICITY: Highly expressed in the heart, moderately in  
 the lung and placenta and at low levels in the liver, skeletal  
 muscle, kidney, pancreas, spleen, lymph node, thymus, bone marrow  
 and fetal liver. No expression was seen in adult brain or  
 peripheral blood leukocytes.  
 CC PTM: Synthesized in the endoplasmic reticulum as an inactive form  
 which is proteolytically cleaved by a furin-like convertase in the  
 trans-Golgi network before it reaches the plasma membrane to yield  
 an active, ligand-accessible form. Cleavage results in a C-  
 terminal fragment N(TM) and a N-terminal fragment N(EC). Following  
 ligand binding, it is cleaved by TNF-alpha converting enzyme  
 (TACE) to yield a membrane-associated intermediate fragment called  
 notch extracellular truncation (NEXT). This fragment is then  
 cleaved by presenilin dependent gamma-secretase to release a  
 notch-derived peptide containing the intracellular domain (NICD)  
 from the membrane (By similarity).  
 CC -!- PTM: Phosphorylation (By similarity).  
 CC PTM: The poly-Lys region of NOTCH4 (in the signal  
 peptide) is polymorphic and the number of Lys varies in the  
 population (From 6 to 12).  
 CC -!- SIMILARITY: Belongs to the NOTCH family.  
 CC -!- SIMILARITY: Contains 28 EGF-like domains.  
 CC -!- SIMILARITY: Contains 3 Lnn/Notch repeats.  
 CC -!- SIMILARITY: Contains 5 ANK repeats.  
 CC -!- POSITION: Ref.1 sequence differs from that shown due to frameshift  
 in position 1438 to 1433.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to license@isb-sib.ch).

CC	DR	-----
	EMBL; DR	D63395; BA09708.1; ALT FRAME.
	EMBL; DR	D86566; BA13116.1; -.
	EMBL; DR	U8529; AAC3288.1; -.
	EMBL; DR	U89335; AAC63097.1; -.
	EMBL; DR	AB02361; BAB2017.1; -.
	EMBL; DR	AB024520; BAA8951.1; -.
	EMBL; DR	AB024578; BAA8852.1; -.
	DR	HSSP; P08709; 1B99.
	DR	Genew; HGNC; 7884; NOTCH4.
	DR	MIM; 164951; -.
	DR	InterPro; IPR002110; ANK.
	DR	InterPro; IPR00152; Asx_hydroxyl_S.
	DR	InterPro; IPR00742; EGF-2.
	DR	InterPro; IPR01438; EGF-CA.
	DR	InterPro; IPR01438; EGF-CA.
	DR	InterPro; IPR06209; EGF-like.
	DR	InterPro; IPR02049; Laminin_EGF.
	DR	InterPro; IPR008297; Notch.
	DR	InterPro; IPR000800; Notch_dom.
	DR	Pfam; PF00023; ank; 6.
	DR	Pfam; PF00008; EGF; 26.
	DR	Pfam; PF00066; notch; 2.
	DR	PIRSF; PIREF002279; Notch; 1.
	DR	PRINTS; PRO0010; EGFBLOOD.
	DR	PRINTS; PRO0011; EGFLAMININ.
	DR	PRINTS; PRO1452; NOTCH.
	DR	SMART; SM00248; ANK; 5.
	DR	SMART; SM00179; EGF-CA; 11.
	DR	SMART; SM00044; NL; 3.
	DR	PROSITE; PS00187; EGF-CA; 9.
	DR	PROSITE; PS0026; EGF; 3; 28.
	DR	PROSITE; PS00088; ANK_REPEAT; 5.
	DR	PROSITE; PS00010; ASX_HYDROXYL; 11.
	DR	PROSITE; PS00022; EGF-1; 28.
	DR	PROSITE; PS01186; EGF-2; 21.
	FT	CHAIN 1 23
	FT	CHAIN 2 2003
	FT	CHAIN 1432 2003
	FT	CHAIN 1467 2003
	FT	DOMAIN 24 1447
	FT	TRANSMEM 1448 1468
	FT	DOMAIN 1469 2003
	FT	DOMAIN 24 63
	FT	DOMAIN 64 115
	FT	DOMAIN 156 192
	FT	DOMAIN 194 232
	FT	DOMAIN 234 274
	FT	DOMAIN 276 312
	FT	DOMAIN 314 353
	FT	DOMAIN 355 391
	FT	DOMAIN 392 430
	FT	DOMAIN 432 473
	FT	DOMAIN 475 511
	FT	DOMAIN 513 549
	FT	DOMAIN 551 587
	FT	DOMAIN 589 625
	FT	DOMAIN 626 659
	FT	DOMAIN 661 689
	FT	DOMAIN 691 727
	FT	DOMAIN 729 765
	FT	DOMAIN 767 803
	FT	DOMAIN 806 842
	FT	DOMAIN 844 880
	FT	DOMAIN 882 928
	CC	-----
	CC	NOTCH_INTRACELLULAR_DOMAIN (BY SIMILARITY); NOTCH_EXTRACELLULAR_TRUNCATION (BY SIMILARITY); NOTCH_INTRACELLULAR_DOMAIN (BY SIMILARITY); EXTRACELLULAR (POTENTIAL). POTENTIAL. CYTOPLASMIC (POTENTIAL). NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 4.

FT	DOMAIN	930	966	EGF-LIKE 24.	CC	-!- SIMILARITY: Contains 36 EGF-like domains.
FT	DOMAIN	968	1004	EGF-LIKE 25.	CC	-!- SIMILARITY: Contains 6 ANK repeats.
FT	DOMAIN	1006	1044	EGF-LIKE 25.	CC	-!- SIMILARITY: Contains 6 ANK repeats.
FT	DOMAIN	1046	1085	EGF-LIKE 26.	CC	-!- SIMILARITY: Contains 6 ANK repeats.
FT	DOMAIN	1087	1126	EGF-LIKE 27.	CC	-!- SIMILARITY: Contains 6 ANK repeats.
FT	DOMAIN	1130	1171	EGF-LIKE 28.	CC	-!- SIMILARITY: Contains 6 ANK repeats.
FT	DOMAIN	1472	1475	POLY-ARG.	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
FT	REPEAT	1165	1212	LIN/NOTCH 1.	CC	use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
FT	REPEAT	1213	1245	LIN/NOTCH 2.	CC	use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
FT	REPEAT	1247	1286	LIN/NOTCH 3.	CC	use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
FT	REPEAT	1663	1698	ANK 1.	CC	use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
FT	REPEAT	1666	1698	ANK 2.	CC	use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
FT	REPEAT	1700	1732	ANK 3.	CC	use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
FT	REPEAT	1733	1765	ANK 4.	CC	use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
FT	REPEAT	1766	1798	ANK 5.	CC	use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
FT	DISULFD	28	41	BY SIMILARITY.	DR	InterPro; IPR001522; AxS_hydroxyl_S.
FT	DISULFD	35	51	BY SIMILARITY.	DR	InterPro; IPR001742; EGF_2.
FT	DISULFD	53	62	BY SIMILARITY.	DR	InterPro; IPR001881; EGF_Ca.
FT	DISULFD	68	80	BY SIMILARITY.	DR	InterPro; IPR001438; EGF_II.
FT	DISULFD	74	103	BY SIMILARITY.	DR	InterPro; IPR005209; EGF_Like.
FT	DISULFD	105	114	BY SIMILARITY.	DR	InterPro; IPR002049; Laminin_EGF.
FT	DISULFD	122	133	BY SIMILARITY.	DR	InterPro; IPR008297; Notch.
FT	DISULFD	127	143	BY SIMILARITY.	DR	InterPro; IPR00800; Notch_dom.
FT	DISULFD	145	154	BY SIMILARITY.	DR	Pfam; PF0023; ank; 6.
FT	DISULFD	160	171	BY SIMILARITY.	DR	Pfam; PF0008; EGF; 36.
FT	DISULFD	165	180	BY SIMILARITY.	DR	Pfam; PF0066; notch; 3.
FT	DISULFD	182	191	BY SIMILARITY.	DR	Pfam; PF002279; Notch; 1.
FT	DISULFD	198	211	BY SIMILARITY.	DR	PRINTS; PRO001; EGFBLOOD.
Query Match		3.1%	Score 151.5;	DB 1;	Length 2003;	
Best Local Similarity		27.9%	Pred. No. 0.36;	Mismatches 12;	Indels 30;	
Matches		39;	Conservative		Gaps 7;	
QY	434	PVNC---AEGPGSMWKGDCDKACNNACDWDGGDCSGNSGGSRYIAGGGGTSIGV	489	DR	SMART; SMC0004; NL; 2.	
Db	1166	P P C Q K P G K G C E G ---S G D A C G S P G C N W D G C	1207	DR	PROSITE; PS0088; ANK_REPEAT; 4.	
QY	490	GH W Q F G G G I N S V S Y C --- --- --- --- --- --- --- --- ---	518	DR	PROSITE; PS0022; EGF_1; 34.	
Db	1208	P P W K --- G C P S H R C W L F R Q G Q C H Q C D S E C L F G D C B A P T P A C T P A Y D Q Y C H H F I	1264	DR	PROSITE; PS50026; EGF_3; 36.	
QY	519	-- C D Q A C N V L S C G F D A C	535	DR	PROSITE; PS01187; EGF_Ca; 21.	
Db	1265	N G C E R G G N T A E C G W G D G C	1284	DR	Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain; KW Transmembrane; Signal; Glycoprotein; SIGNAL; 19	
RESULT 11				FT	CHAIN	
NOTC_XENLA				FT	DOMAIN	20
ID	NOTC_XENLA			FT	TRANSMEM	1729
AC	P21783;	STANDARD;	PRT; 2524 AA.	FT	DOMAIN	1751
DT	01-OCT-1991 (Rel. 18, Created)			FT	DOMAIN	20
DT	01-OCT-1996 (Rel. 18, Last sequence update)			FT	DOMAIN	57
DE	28-FEB-2003 (Rel. 41, Last annotation update)			FT	DOMAIN	58
NEurogenic locus notch protein homolog precursor (XOTCH protein).	XOTCH.			FT	DOMAIN	99
OS	Xenopus laevis (African clawed frog).			FT	DOMAIN	102
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batracia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopoda; Xenopodinae; Xenopus.			FT	DOMAIN	140
[1]	NCBI_TaxID:8355;			FT	DOMAIN	179
RP	SEQUENCE FROM N.A.			FT	DOMAIN	217
RX	MEDLINE=9038285; PubMed=2402639;			FT	DOMAIN	254
RA	Coffman C., Harris W., Kintner C.;			FT	DOMAIN	292
RT	"Xotch", the Xenopus homolog of Drosophila notch. n;			FT	DOMAIN	332
RL	Science 249:1438-1441(1990).			FT	DOMAIN	370
RN	[2]			FT	DOMAIN	409
RP	REVISONS TO 1759-1782.			FT	DOMAIN	449
RA	Kintner C.:			FT	DOMAIN	487
RX	Submitted (UNP-1996) to the EMBL/GenBank/DBJ databases.			FT	DOMAIN	525
CC	-- SUBCELLULAR LOCATION: Type I membrane protein.			FT	DOMAIN	563
-- DEVELOPMENTAL STAGE: Expressed almost uniformly in early embryos.				FT	DOMAIN	600
CC	-- SIMILARITY: Belongs to the NOTCH family.			FT	DOMAIN	638
				FT	DOMAIN	675
				FT	DOMAIN	713
				FT	DOMAIN	750
				FT	DOMAIN	788
				FT	DOMAIN	826
				FT	DOMAIN	866
				FT	DOMAIN	904
				FT	DOMAIN	942
				FT	DOMAIN	980
				FT	DOMAIN	982
				FT	DOMAIN	1018

FT	DOMAIN	1020	1056	EGF-LIKE 27. CALCIUM-BINDING (POTENTIAL).	FT	DISULFID	719	729	BY SIMILARITY.
FT	DOMAIN	1058	1094	EGF-LIKE 28.	FT	DISULFID	724	738	BY SIMILARITY.
FT	DOMAIN	1096	1142	EGF-LIKE 29.	FT	DISULFID	740	749	BY SIMILARITY.
FT	DOMAIN	1144	1180	EGF-LIKE 30. CALCIUM-BINDING (POTENTIAL).	FT	DISULFID	756	767	BY SIMILARITY.
FT	DOMAIN	1182	1220	EGF-LIKE 31. CALCIUM-BINDING (POTENTIAL).	FT	DISULFID	761	776	BY SIMILARITY.
FT	DOMAIN	1226	1264	EGF-LIKE 32. CALCIUM-BINDING (POTENTIAL).	FT	DISULFID	778	787	BY SIMILARITY.
FT	DOMAIN	1266	1304	EGF-LIKE 33.	FT	DISULFID	794	805	BY SIMILARITY.
FT	DOMAIN	1306	1346	EGF-LIKE 34.	FT	DISULFID	799	814	BY SIMILARITY.
FT	DOMAIN	1347	1383	EGF-LIKE 35.	FT	DISULFID	816	825	BY SIMILARITY.
FT	REPEAT	1386	1424	EGF-LIKE 36.	FT	DISULFID	832	843	BY SIMILARITY.
FT	REPEAT	1424	1478	LIN/NOTCH 1.	FT	DISULFID	837	854	BY SIMILARITY.
FT	REPEAT	1479	1520	LIN/NOTCH 2.	FT	DISULFID	856	865	BY SIMILARITY.
FT	REPEAT	1521	1560	LIN/NOTCH 3.	FT	DISULFID	872	883	BY SIMILARITY.
FT	REPEAT	1876	1919	ANK 1.	FT	DISULFID	877	892	BY SIMILARITY.
FT	REPEAT	1924	1953	ANK 2.	FT	DISULFID	894	903	BY SIMILARITY.
FT	REPEAT	1957	1987	ANK 3.	FT	DISULFID	910	921	BY SIMILARITY.
FT	REPEAT	1991	2020	ANK 4.	FT	DISULFID	915	930	BY SIMILARITY.
FT	REPEAT	2024	2053	ANK 5.	FT	DISULFID	932	941	BY SIMILARITY.
FT	DISULFID	2057	2086	ANK 6.	FT	DISULFID	986	997	BY SIMILARITY.
FT	DISULFID	22	35	BY SIMILARITY.	FT	DISULFID	991	1006	BY SIMILARITY.
FT	DISULFID	29	45	BY SIMILARITY.	FT	DISULFID	1008	1017	BY SIMILARITY.
FT	DISULFID	47	55	BY SIMILARITY.	FT	DISULFID	1024	1035	BY SIMILARITY.
FT	DISULFID	62	74	BY SIMILARITY.	FT	DISULFID	1029	1044	BY SIMILARITY.
FT	DISULFID	68	87	BY SIMILARITY.	FT	DISULFID	1046	1055	BY SIMILARITY.
FT	DISULFID	89	98	BY SIMILARITY.	FT	DISULFID	1062	1073	BY SIMILARITY.
FT	DISULFID	106	117	BY SIMILARITY.	FT	DISULFID	1067	1082	BY SIMILARITY.
FT	DISULFID	111	128	BY SIMILARITY.	FT	DISULFID	1084	1093	BY SIMILARITY.
FT	DISULFID	130	139	BY SIMILARITY.	FT	DISULFID	1121	1129	BY SIMILARITY.
FT	DISULFID	145	156	BY SIMILARITY.	FT	DISULFID	1130	1130	BY SIMILARITY.
FT	DISULFID	150	165	BY SIMILARITY.					
FT	DISULFID	167	176	BY SIMILARITY.					
FT	DISULFID	183	194	BY SIMILARITY.					
FT	DISULFID	188	203	BY SIMILARITY.					
FT	DISULFID	205	214	BY SIMILARITY.					
FT	DISULFID	221	232	BY SIMILARITY.					
FT	DISULFID	226	242	BY SIMILARITY.					
FT	DISULFID	244	253	BY SIMILARITY.					
FT	DISULFID	260	271	BY SIMILARITY.					
FT	DISULFID	265	280	BY SIMILARITY.					
FT	DISULFID	282	291	BY SIMILARITY.					
FT	DISULFID	298	311	BY SIMILARITY.					
FT	DISULFID	305	320	BY SIMILARITY.					
FT	DISULFID	322	331	BY SIMILARITY.					
FT	DISULFID	338	349	BY SIMILARITY.					
FT	DISULFID	343	358	BY SIMILARITY.					
FT	DISULFID	360	369	BY SIMILARITY.					
FT	DISULFID	375	386	BY SIMILARITY.					
FT	DISULFID	380	397	BY SIMILARITY.					
FT	DISULFID	399	408	BY SIMILARITY.					
FT	DISULFID	415	428	BY SIMILARITY.					
FT	DISULFID	422	437	BY SIMILARITY.					
FT	DISULFID	439	448	BY SIMILARITY.					
FT	DISULFID	455	466	BY SIMILARITY.					
FT	DISULFID	460	475	BY SIMILARITY.					
FT	DISULFID	477	486	BY SIMILARITY.					
FT	DISULFID	493	504	BY SIMILARITY.					
FT	DISULFID	509	513	BY SIMILARITY.					
FT	DISULFID	515	524	BY SIMILARITY.					
FT	DISULFID	531	542	BY SIMILARITY.					
FT	DISULFID	536	551	BY SIMILARITY.					
FT	DISULFID	553	562	BY SIMILARITY.					
FT	DISULFID	569	579	BY SIMILARITY.					
FT	DISULFID	574	588	BY SIMILARITY.					
FT	DISULFID	590	599	BY SIMILARITY.					
FT	DISULFID	606	617	BY SIMILARITY.					
FT	DISULFID	611	626	BY SIMILARITY.					
FT	DISULFID	628	637	BY SIMILARITY.					
FT	DISULFID	644	654	BY SIMILARITY.					
FT	DISULFID	649	663	BY SIMILARITY.					
FT	DISULFID	665	674	BY SIMILARITY.					
FT	DISULFID	692	701	BY SIMILARITY.					
FT	DISULFID	686	701	BY SIMILARITY.					
FT	DISULFID	703	712	BY SIMILARITY.					

Query Match 3.0%; Score 149.5; DB 1; Length 2524;  
Best Local Similarity 19.6%; Pred. No. 0.67; Mismatches 217; Indels 275; Gaps 42;

Matches 143; Conservative 95; Mismatches 143; Indels 275; Gaps 42;

Qy 412 GKDWWPPDPYSHKGKQYLTWPMVPAEGCGSWKDG-YDKACNNASADWDGDCSG 470

Db 1435 GKNITPPDNIDCNEQ-----LADNKVNANCNNHAGNCWGDGDC-- 1477

Qy 471 NSGGSRVYIAGGGGTGSIGVGHPMQ-----EGGINSVSYCNG-GCA----- 510

Db 1478 -----SLNFNDPWFCNCTOSLOCWKFENDG-KCDSDQCNNTGCLYDGFDCQR 1521

Qy 511 -----NSWLADKF---CDQACNVLSCGFDAGDGCGO-DHFH--LYKVLIP--- 550

Db 1522 VEVQCNPLVDQYCDDHFQDGHDQCGNNACEWGDGCANMPENLAEGTLVILWMPPE 1581

Qy 551 ---NOTHY-----IPKEE-CLPYF-SPAEYAKRGVEGA--YSNDP-- 584

Db 1582 LKRNNSVNFRELRSVLHNVVKDKSOGEKYKIPYGNBEEELKKHHRSTDWPSA 1641

Qy 585 -----IIRHASTANK--WKTIIHIMMSGMNAITHRNLT 617

Db 1642 IFSIWKEISILGRRRELDMEVRGSIYLEIDRQKTSQQCFNATDAFLGALAS 1701

Qy 618 QNTDE-EPKMQITVEVTREGKLNSTAQKVENYSPITL-----PEA 662

Db 1702 LGSLHLLTSLYKIE-AVKSENMETPK-PSTLYPMUMLVPLLIIFVMMIVNKGRREHG 1759

Qy 663 EILED--IPKEEPRPKGRHD-----VNST 686

Db 1760 OLWPFDPGFLPKPER--SKCKRDRIGEDSVGLKPIKNTGDSFMDDNQENGDBETLENKR 1817

Qy 687 RRAQEBVKIP-LVNLSLIPKD-----AQLSINTL-----DIQLEHGDITLG-- 727

Db 1818 FRPEEQLVILPEJLDDKTPROMTRQHDAADRISMAAPTPOGEREADCNDVNGPDG 1877

Qy 728 -----XNLSKALLRSLFMMQSCHKTKNQAITDTENDSIVAPO 766

Db 1878 PTPLIMASCGGGLELTGNSSEEDASAMWISPI-CQGAQHLNQDRTGETALHARY 1935

Qy 767 EK-QVHKSTLPNS--LGYSERQRLTPAVSKUNGRDQ-----QNPREDLT---- 812

Db 1936 ARADAQKLUELESSADANVQDNNGRTPLHAA--VAADAQGVQILIRNRTDADMFDG 1992

CC (TACE) to yield a membrane-associated intermediate fragment called  
 CC notch extracellular truncation (NEXT). This fragment is then  
 CC cleaved by presenilin dependent gamma-secretase to release a  
 CC notch-derived peptide containing the intracellular domain (NICD)  
 CC from the membrane (BY similarity).  
 CC -!- PTM: Phosphorylated (BY similarity).  
 CC -!- SIMILARITY: Belongs to the NOTCH family.  
 CC -!- SIMILARITY: Contains 34 EGF-like domains.  
 CC -!- SIMILARITY: Contains 5 ANK repeats.  
 CC -!- SIMILARITY: Contains 3 Lin/Notch repeats.  
 RESULT 12  
 NTC3 RAT STANDARD; PRT; 2319 AA.  
 ID NTC3 RAT  
 AC Q9R177;  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DB Neurogenic locus notch homolog protein 3 precursor (Notch 3).  
 OS Rattus norvegicus (Rat).  
 OC Bokaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TAXID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Haritunians T., Boulter J., Weimaster G., Schanen N.C.;  
 RT "Rattus norvegicus mRNA for Notch 3.";  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP FUNCTION:  
 RX MEDLINE=21094508; PubMed=11182080;  
 RA Tanigaki K., Nogaki F., Takahashi J., Tashiro K., Kurooka H.,  
 RA Honjo T.;  
 RT "Notch1 and Notch3 instructively restrict bFGF-responsive multipotent  
 neural progenitor cells to an astroglial fate.";  
 RL Neuron 29:45-55(2001).  
 RN [3]  
 TISSUE SPECIFICITY:  
 RT MEDLINE=21331789; PubMed=11438922;  
 RA Irvin D.K., Zurcher S.D., Nguyen T., Weimaster G., Kornblum H.I.;  
 RT "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple  
 functional roles for the Notch-DSL signaling system during brain  
 development.";  
 J. Comp. Neurol. 436:167-181 (2001).  
 CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands  
 CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.  
 CC Upon ligand activation through the released notch intracellular  
 CC domain (NICD) it forms a transcriptional activator complex with  
 CC RBP-J kappa and activates genes of the enhancer of split locus.  
 CC Affects the implementation of differentiation, proliferation and  
 CC apoptotic programs (BY similarity). Acts instructively to control  
 CC the cell fate determination of CNS multipotent progenitor cells,  
 CC resulting in astrogial induction and neuron/oligodendrocyte  
 CC suppression.  
 CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(<sup>TM</sup>) and a N-  
 CC terminal fragment N(EC), which are probably linked by disulfide  
 CC bonds (BY similarity).  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following  
 CC proteolytical processing NICD is translocated to the nucleus.  
 CC TISSUE SPECIFICITY: Expressed in postnatal central nervous system  
 CC (CNS) germinal zones and, in early postnatal life, within  
 CC numerous cells throughout the CNS. It is more highly localized  
 CC to ventricular germinal zones.  
 CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form  
 CC which is proteolytically cleaved by a furin-like convertase in the  
 CC trans-Golgi network before it reaches the plasma membrane to yield  
 CC an active, ligand-accessible form. Cleavage results in a C-  
 CC terminal fragment N(<sup>TM</sup>) and a N-terminal fragment N(EC). Following  
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme

		Matches	36;	Conservative	12;	Mismatches	35;	Indels	56;	Gaps
FT DOMAIN	584	620								
FT DOMAIN	622	657	EGF-LIKE 15,	CALCIUM-BINDING (POTENTIAL).						
FT DOMAIN	659	695	EGF-LIKE 16,	CALCIUM-BINDING (POTENTIAL).						
FT DOMAIN	697	732	EGF-LIKE 17,	CALCIUM-BINDING (POTENTIAL).						
FT DOMAIN	736	772	EGF-LIKE 18.							
FT DOMAIN	773	810	EGF-LIKE 19.							
FT DOMAIN	812	849	EGF-LIKE 20.							
FT DOMAIN	851	887	EGF-LIKE 21,	CALCIUM-BINDING (POTENTIAL).						
FT DOMAIN	885	924	EGF-LIKE 22,	CALCIUM-BINDING (POTENTIAL).						
FT DOMAIN	926	962	EGF-LIKE 23,	CALCIUM-BINDING (POTENTIAL).						
FT DOMAIN	964	1000	EGF-LIKE 24.							
FT DOMAIN	1036	1084	EGF-LIKE 25.							
FT DOMAIN	1122	1160	EGF-LIKE 26.							
FT DOMAIN	1124	1162	EGF-LIKE 27.							
FT DOMAIN	1205	1246	EGF-LIKE 28.							
FT DOMAIN	1247	1289	EGF-LIKE 29,	CALCIUM-BINDING (POTENTIAL).						
FT DOMAIN	1248	1291	EGF-LIKE 30,	CALCIUM-BINDING (POTENTIAL).						
FT DOMAIN	1337	1327	EGF-LIKE 31.							
FT REPEAT	1384	1375	EGF-LIKE 32.							
FT REPEAT	1425	1420	EGF-LIKE 33.							
FT REPEAT	1461	1459	EGF-LIKE 34.							
FT REPEAT	1469	1503	LIN/NOTCH 1.							
FT REPEAT	1840	1869	LIN/NOTCH 2.							
FT REPEAT	1873	1903	LIN/NOTCH 3.							
FT REPEAT	1907	1936	ANK 1.							
FT REPEAT	1940	1969	ANK 2.							
FT REPEAT	1973	2002	ANK 3.							
FT SITE	1574	1574	ANK 4.							
FT DISULFID	44	56	ANK 5.							
FT DISULFID	67	69	CLEAVAGE (BY FURIN-LIKE PROTEASE) (BY SIMILARITY).							
FT DISULFID	78	84	BY SIMILARITY.							
FT DISULFID	95	98	BY SIMILARITY.							
FT DISULFID	108	108	BY SIMILARITY.							
FT DISULFID	110	119	BY SIMILARITY.							
FT DISULFID	135	135	BY SIMILARITY.							
FT DISULFID	146	146	BY SIMILARITY.							
FT DISULFID	148	157	BY SIMILARITY.							
FT DISULFID	174	175	BY SIMILARITY.							
FT DISULFID	179	185	BY SIMILARITY.							
FT DISULFID	187	196	BY SIMILARITY.							
FT DISULFID	203	214	BY SIMILARITY.							
FT DISULFID	208	224	BY SIMILARITY.							
FT DISULFID	226	235	BY SIMILARITY.							
FT DISULFID	242	253	BY SIMILARITY.							
FT DISULFID	247	262	BY SIMILARITY.							
FT DISULFID	264	273	BY SIMILARITY.							
FT DISULFID	280	293	BY SIMILARITY.							
FT DISULFID	287	302	BY SIMILARITY.							
FT DISULFID	304	313	BY SIMILARITY.							
FT DISULFID	320	331	BY SIMILARITY.							
FT DISULFID	325	340	BY SIMILARITY.							
FT DISULFID	342	351	BY SIMILARITY.							
FT DISULFID	357	368	BY SIMILARITY.							
FT DISULFID	362	379	BY SIMILARITY.							
FT DISULFID	390	390	BY SIMILARITY.							
FT DISULFID	391	410	BY SIMILARITY.							
FT DISULFID	419	419	BY SIMILARITY.							
FT DISULFID	421	430	BY SIMILARITY.							
FT DISULFID	437	448	BY SIMILARITY.							
FT DISULFID	442	457	BY SIMILARITY.							
FT DISULFID	459	468	BY SIMILARITY.							
FT DISULFID	475	486	BY SIMILARITY.							
FT DISULFID	480	495	BY SIMILARITY.							
FT DISULFID	497	506	BY SIMILARITY.							
FT DISULFID	513	524	BY SIMILARITY.							
FT DISULFID	518	533	BY SIMILARITY.							
FT DISULFID	534	544	BY SIMILARITY.							
FT DISULFID	551	561	BY SIMILARITY.							
Query Match Best Local Similarity	3.0%;	Score 148; DB 1; Length 2319;	QY	436	FNCAGCPGWSKTDGKDACKNSACWDGGCGSGNSGSSRYIAGGGTGSAIVGVHFWQ-	494				
	pred. No. 0.73;		Db	1387	PRCPRACQAQRGQDQNCRECSPGCWGCGC-----SLANVDPWRO	142				
			QY	495	--RGGGINSVSYCCN---QGCANSWLADKFC	519				
			Db	1430	CEALQCWRLFNNSRCPACSSPACYDNFDCYSGGRDT-CNPVVKYCADHF-ADGRC	148				
			QY	520	DQACNVISCGFRAGDGCQD	538				
			Db	1487	DOCNTBECGMDGLDCASE	1505				
			RESULT	13						
			ID	NTC3_MOUSE						
			STANDARD:							
			FRT:	2318 AA.						
			AC	Q61982;						
			DT	OII-NON-1997 (Rel. 35, Created)						
			DT	OII-NON-1997 (Rel. 35, Last sequence update)						
			DT	28-FEB-2003 (Rel. 41, Last annotation update)						
			DB	Neurogenic locus notch homolog protein 3 precursor (Notch 3).						
			GN	NOTCH3.						
			OS	Mus musculus (Mouse).						
			OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Buteraria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.						
			OX	NCBI_TaxID:10090;						
			RN	[1]						
			RP	SEQUENCE FROM N.A.						
			RC	STRAIN=ICR X Swiss Webster;						
			RX	MEDLINE=9301556; PubMed=7918097;						
			RA	Lindell M., Dalstrand J., Landahl U.;						
			RT	"The novel Notch homologue mouse Notch 3 lacks specific epidermal growth factor-repeats and is expressed in proliferating neuroepithelium"						
			RT	Nech. Dev. 46:123-136(1994).						
			RL	[12]						
			RP	POST-TRANSLATIONAL PROCESSING AND MUTAGENESIS OF MET-1664.						
			RX	MEDLINE=2134376; PubMed=1149941;						
			RA	Nizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;						
			RT	"Conservation of the biochemical mechanisms of signal transduction among mammalian Notch family members."						
			RL	Proc. Natl. Acad. Sci. U.S.A. 98:3026-3031(2001).						
			CC	-!- FUNCTION: Functions as a receptor for membrane-bound ligands Jagged1, Jagged2 and Delta1 to regulate cell-cell fate determination.						
			CC	Upon ligand activation through the released notch intracellular domain (NICD) it forms a transcriptional activator complex with RBP-J kappa and activates genes of the enhancer of split locus.						
			CC	Affects the implementation of differentiation, proliferation and apoptotic programs (By similarity). May play a role during CNS development.						
			CC	-!- SUBUNIT: Heterodimer of a C-terminal fragment N( <sup>TM</sup> ) and a N-terminal fragment NICD which are probably linked by disulfide bonds.						
			CC	--SUBCELLULAR LOCATION: Type I membrane protein. Following proteolytical processing NICD is translocated to the nucleus.						
			CC	--TISSUE SPECIFICITY: Proliferating neuroepithelium.						
			CC	--DEVELOPMENTAL STAGE: CNS development.						
			CC	--PTM: Synthesized in the endoplasmic reticulum as an inactive form which is proteolytically cleaved by a furin-like convertase to yield an active, ligand-accessible form. Cleavage results in a C-terminal fragment NICD and a N-terminal fragment NICD. Following ligand binding, it is cleaved by TNF-alpha converting enzyme (TACE) to yield a membrane-associated intermediate fragment called						

CC	notch extracellular truncation (NEXT). This fragment is then cleaved by presenilin dependent gamma-secretase to release a notch-derived peptide containing the intracellular domain (NICD) from the membrane.	FT DOMAIN 470	FT DOMAIN 506	EGR-LIKE 12, CALCIUM-BINDING (POTENTIAL).
CC	-!- PPM: Phosphorylated.	FT DOMAIN 508	FT DOMAIN 544	EGR-LIKE 13, CALCIUM-BINDING (POTENTIAL).
CC	-!- SIMILARITY: Belongs to the NOTCH family.	FT DOMAIN 546	FT DOMAIN 583	EGR-LIKE 14, CALCIUM-BINDING (POTENTIAL).
CC	-!- SIMILARITY: Contains 34 EGF-like domains.	FT DOMAIN 583	FT DOMAIN 619	EGR-LIKE 15, CALCIUM-BINDING (POTENTIAL).
CC	-!- SIMILARITY: Contains 3 Lnn/Notch repeats.	FT DOMAIN 621	FT DOMAIN 656	EGR-LIKE 16, CALCIUM-BINDING (POTENTIAL).
CC	-----	FT DOMAIN 658	FT DOMAIN 694	EGR-LIKE 17, CALCIUM-BINDING (POTENTIAL).
CC	-----	FT DOMAIN 696	FT DOMAIN 731	EGR-LIKE 18.
CC	-----	FT DOMAIN 735	FT DOMAIN 771	EGR-LIKE 19.
CC	-----	FT DOMAIN 772	FT DOMAIN 809	EGR-LIKE 20.
CC	-----	FT DOMAIN 811	FT DOMAIN 848	EGR-LIKE 21, CALCIUM-BINDING (POTENTIAL).
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch">http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch</a> ).	FT DOMAIN 850	FT DOMAIN 886	EGR-LIKE 22, CALCIUM-BINDING (POTENTIAL).
CC	-----	FT DOMAIN 888	FT DOMAIN 923	EGR-LIKE 23, CALCIUM-BINDING (POTENTIAL).
CC	-----	FT DOMAIN 925	FT DOMAIN 959	EGR-LIKE 24.
CC	-----	FT DOMAIN 963	FT DOMAIN 1001	EGR-LIKE 25.
CC	-----	FT DOMAIN 1001	FT DOMAIN 1035	EGR-LIKE 26.
CC	-----	FT DOMAIN 1035	FT DOMAIN 1083	EGR-LIKE 27.
CC	-----	FT DOMAIN 1083	FT DOMAIN 1121	EGR-LIKE 28.
CC	-----	FT DOMAIN 1121	FT DOMAIN 1159	EGR-LIKE 29, CALCIUM-BINDING (POTENTIAL).
CC	-----	FT DOMAIN 1159	FT DOMAIN 1204	EGR-LIKE 30, CALCIUM-BINDING (POTENTIAL).
DR	InterPro; IPR002110; ANK.	FT DOMAIN 1204	FT DOMAIN 1245	EGR-LIKE 31.
DR	InterPro; IPR001512; ASX hydroxyl_S.	FT DOMAIN 1245	FT DOMAIN 1288	EGR-LIKE 32.
DR	InterPro; IPR007742; EGF 2.	FT DOMAIN 1288	FT DOMAIN 1326	EGR-LIKE 33.
DR	InterPro; IPR0040; IEDM.	FT DOMAIN 1326	FT DOMAIN 1374	EGR-LIKE 34.
DR	HSSP; P0040; IEDM.	FT DOMAIN 1374	FT DOMAIN 1428	LIN/NOTCH 1.
DR	GO: 0005515; F-protein binding; IPI.	FT DOMAIN 1428	FT DOMAIN 1467	LIN/NOTCH 2.
DR	GO: 0007219; P:N signaling pathway; IC.	FT DOMAIN 1467	FT DOMAIN 1503	LIN/NOTCH 3.
DR	InterPro; IPR002110; ANK.	FT DOMAIN 1503	FT DOMAIN 1868	ANK 1.
DR	InterPro; IPR007742; EGF 2.	FT DOMAIN 1868	FT DOMAIN 1902	ANK 2.
DR	InterPro; IPR001512; ASX hydroxyl_S.	FT DOMAIN 1902	FT DOMAIN 1935	ANK 3.
DR	InterPro; IPR002049; laminin_EGF.	FT DOMAIN 1935	FT DOMAIN 1968	ANK 4.
DR	InterPro; IPR008297; Notch.	FT DOMAIN 1968	FT DOMAIN 2001	ANK 5.
DR	InterPro; IPR00800; Notch_dom.	FT DOMAIN 2001	FT DOMAIN 1572	CLEAVAGE (BY FURIN-LIKE PROTEASE) (BY SITE)
DR	PRINTS; PR00023; ank; 6.	FT DISULFID 43	FT DISULFID 55	SIMILARITY).
DR	PRINTS; PR00008; EGF; 33.	FT DISULFID 49	FT DISULFID 66	BY SIMILARITY.
DR	PRINTS; PR00066; notch; 3.	FT DISULFID 68	FT DISULFID 77	BY SIMILARITY.
DR	PRINTS; PR0002279; Notch; 1.	FT DISULFID 83	FT DISULFID 94	BY SIMILARITY.
DR	PRINTS; PR00010; EGF_BLOOD.	FT DISULFID 88	FT DISULFID 107	BY SIMILARITY.
DR	PRINTS; PR00011; EGF_LAMININ.	FT DISULFID 109	FT DISULFID 118	BY SIMILARITY.
DR	PRINTS; PR01452; NOTCH.	FT DISULFID 124	FT DISULFID 135	BY SIMILARITY.
DR	SMART; SMC0248; ANK; 6.	FT DISULFID 129	FT DISULFID 145	BY SIMILARITY.
DR	SMART; SMC0119; EGF_Ca; 19.	FT DISULFID 147	FT DISULFID 156	BY SIMILARITY.
DR	SMART; SM0004; NL; 3.	FT DISULFID 163	FT DISULFID 175	BY SIMILARITY.
DR	PROSITE; PS50297; ANK REP REGION; 1.	FT DISULFID 169	FT DISULFID 184	BY SIMILARITY.
DR	PROSITE; PS50086; ANK REPEAT; 4.	FT DISULFID 186	FT DISULFID 195	BY SIMILARITY.
DR	PROSITE; PS00010; ASX_HYDROXYL; 18.	FT DISULFID 202	FT DISULFID 213	BY SIMILARITY.
DR	PROSITE; PS00022; EGF 1; 33.	FT DISULFID 207	FT DISULFID 223	BY SIMILARITY.
DR	PROSITE; PS01186; EGF_2; 27.	FT DISULFID 225	FT DISULFID 234	BY SIMILARITY.
DR	PROSITE; PS50026; EGF_3; 34.	FT DISULFID 241	FT DISULFID 252	BY SIMILARITY.
DR	PROSITE; PS01187; EGF_Ca; 16.	FT DISULFID 246	FT DISULFID 261	BY SIMILARITY.
KW	Receptor; Transcrip regulation; Activator; Differentiation; Developmental protein; Repeat; ANK repeat; EGF-like domain; Transmembrane; Glycoprotein; Signal; Phosphorylation.	FT DISULFID 263	FT DISULFID 272	BY SIMILARITY.
FT	SIGNAL 1	FT DISULFID 279	FT DISULFID 292	BY SIMILARITY.
FT	CHAIN 40	FT DISULFID 286	FT DISULFID 301	BY SIMILARITY.
FT	CHAIN 1630	FT DISULFID 303	FT DISULFID 312	BY SIMILARITY.
FT	CHAIN 1663	FT DISULFID 319	FT DISULFID 330	BY SIMILARITY.
FT	DOMAIN 40	FT DISULFID 324	FT DISULFID 339	BY SIMILARITY.
FT	DOMAIN 1644	FT DISULFID 341	FT DISULFID 350	BY SIMILARITY.
FT	TRANSMEM 1665	FT DISULFID 346	FT DISULFID 356	BY SIMILARITY.
FT	DOMAIN 2242	FT DISULFID 356	FT DISULFID 367	BY SIMILARITY.
FT	DOMAIN 40	FT DISULFID 361	FT DISULFID 378	BY SIMILARITY.
FT	DOMAIN 79	FT DISULFID 367	FT DISULFID 389	BY SIMILARITY.
FT	DOMAIN 157	FT DISULFID 380	FT DISULFID 398	BY SIMILARITY.
FT	DOMAIN 159	FT DISULFID 380	FT DISULFID 409	BY SIMILARITY.
FT	DOMAIN 198	FT DISULFID 409	FT DISULFID 474	BY SIMILARITY.
FT	DOMAIN 237	FT DISULFID 403	FT DISULFID 479	BY SIMILARITY.
FT	DOMAIN 275	FT DISULFID 429	FT DISULFID 505	BY SIMILARITY.
FT	DOMAIN 315	FT DISULFID 436	FT DISULFID 512	BY SIMILARITY.
FT	DOMAIN 352	FT DISULFID 447	FT DISULFID 523	BY SIMILARITY.
FT	DOMAIN 390	FT DISULFID 456	FT DISULFID 532	BY SIMILARITY.
FT	DOMAIN 430	FT DISULFID 457	FT DISULFID 532	BY SIMILARITY.
FT	DOMAIN 468	FT DISULFID 485	FT DISULFID 532	BY SIMILARITY.
Query Match	3.0%	Score 147;	DB 1;	Length 2318;

Best local Similarity 25.9%; Pred. No. 0.85; Matches 36; Conservative 11; Mismatches 36; Indels 56; Gaps 5; CC  
 QY 436 PNCAGGCAGGCGSWIKDGYCCKACNNACDIDGGCSGSGNSGSRVTTAGGGGTGGSIGVSPWQ- 494 CC  
 Db 1386 PRCPRAAQAKRGDQNCDRECNTPGCGFDGGC-----SINVDDPWRQ 1428 CC  
 QY 495 -----GGGINSVSYN---QGCNSMADKFC 519 CC  
 Db 1429 CEALQCWRFLFNNSRCDPACSSPACIYDNDFFCYSGGRDT--CNPVYBKCADHF-ADGRC 1485 CC  
 QY 520 DQACNVLSGGPAGDGCGD 538 CC  
 Db 1486 DQGENTEREGWDGDCAE 1504 CC  
 RESULT 14 CC  
 NTCL\_HUMAN STANDARD; PRT; 2556 AA. CC  
 AC PA6531; DT 01-NOV-1995 (Rel. 32, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE (Translocation-associated notch protein TAN-1).  
 GN NOTCH1 OR TAN1.  
 OS Homo sapiens (Human).  
 OC Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OC NCBI\_TAXID=9606; OK  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE\_Brain;  
 RA "Mann R.S., Blaumüller C.M., Zagouras P.,"  
 RT "Complete human notch 1 (hnt1) cDNA sequence.",  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE OF 1-2444 FROM N.A.  
 RX MEDLINE=91347357; PubMed=831692;  
 RA Ellisen L.W., Bird J., West D.C., Soriano A.L., Reynolds T.C.,  
 RA Smith S.D., Sklar J.;  
 RT "TAN-1, the human homolog of the *Drosophila* notch gene, is broken by chromosomal translocations in T lymphoblastic neoplasms.";  
 RL Cell 66:649-661(1991).  
 RN [3]  
 RP IDENTIFICATION OF LIGANDS.  
 RX MEDLINE=99108765; PubMed=10079256;  
 RA Gray G.E., Mann R.S., Mittiadis E., Henrique D., Carcangiu M.-L.,  
 RA Banks A., Leiman J., Ward D., Ish-Horowicz D., Artavanis-Tsakonas S.;  
 RT "Human ligands of the Notch receptor.";  
 RL Am. J. Pathol. 154:735-794(1999).  
 RN [4]  
 RP INTERACTION WITH DTX1.  
 RX MEDLINE=98250176; PubMed=9590294;  
 RA Matsuno K., Eastman D., Mittiadis T., Quinn A.M., Carcangiu M.L.,  
 RA Ordentlich P., Kedesch T., Artavanis-Tsakonas S.;  
 RT "Human delta $\epsilon$  is a conserved regulator of Notch signaling.";  
 RL Nat. Genet. 19:74-78(1998).  
 CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands. DR SMART; SM0004; NL; 3  
 CC Jagged1, Jagged2 and Delta to regulate cell fate determination. DR PROSITE; PS50297; ANK REP REGION; 1.  
 CC Upon ligand activation through the released notch intracellular DR PROSITE; PS00088; ANK REPEAT; 4.  
 CC domain (NICD) it forms a transcriptional activator complex with DR PROSITE; PS00110; ASX HYDROXYL; 20.  
 CC RBP-J kappa and activates genes of the enhancer of split locus. DR PROSITE; PS00022; EGF-1; 34.  
 CC Affects the implementation of differentiation, proliferation and DR PROSITE; PS00186; EGF-2; 26.  
 CC apoptotic programs. May be important for normal lymphocyte DR PROSITE; PS50026; EGF-3; 36.  
 CC function. In altered form, may contribute to transformation or DR PROSITE; PS0117; EGF-CA; 18.  
 CC progression in some T-cell neoplasms. Involved in the maturation DR Receptor; Transcription regulation; Activator; Differentiation;  
 CC of both CD4+ and CD8+ cells in the thymus. (By similarity). DR Developmental protein; Repeat; ANK repeat; EGF-like domain;  
 CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N( $\text{TM}$ ) and a N- DR Transmembrane; Glycoprotein; Signal; Phosphorylation.  
 CC terminal fragment N( $\text{EC}$ ) which are probably linked by disulfide FT SIGNAL  
 CC bonds (By similarity). Interacts with DTX1 and DTX2. FT  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following FT  
 CC proteolytical processing NICD is translocated to the nucleus (By CHAIN 175 2556

CC similarity).  
 CC -!- TISSUE SPECIFICITY: In fetal tissues most abundant in spleen, brain stem and lung. Also present in most adult tissues where it is found mainly in lymphoid tissues.  
 CC -!- TM: Synthesized in the endoplasmic reticulum as an inactive form which is proteolytically cleaved by a furin-like convertase in the trans-Golgi network before it reaches the plasma membrane to yield an active, ligand-accessible form. Cleavage results in a C-terminal fragment (NTM) and a N-terminal fragment (NTC). Following ligand binding, it is cleaved by TNF-alpha converting enzyme (TACE) to yield a membrane-associated intermediate fragment called notch extracellular truncation (NEXX). This fragment is then cleaved by presenilin dependent gamma-secretase to release a notch-derived peptide containing the intracellular domain (NICD) from the membrane (By similarity).  
 CC -!- PTM: Phosphorylated (By similarity).  
 CC -!- DISEASE: NOTCH1 truncation is associated with T-cell acute lymphoblastic leukemia.  
 CC -!- SIMILARITY: Belongs to the NOTCH family.  
 CC -!- SIMILARITY: Contains 3 EGF-like domains.  
 CC -!- SIMILARITY: Contains 5 ANK repeats.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).  
 CC  
 DR EML1; AF308602; AG3384; 1; -. DR  
 DR EMBL; MT3980; AA60614; 1; -. DR  
 DR HSSEB; P00740; IEDB.  
 DR Genew; HGNC; 7881; NOTCH1.  
 DR MIM; 190198; -. DR  
 DR GO; GO:0016021; C:integral to membrane; NAS.  
 DR GO; GO:0006355; P:immune response; NAS.  
 DR InterPro; IPR002110; ANK.  
 DR InterPro; IPR00152; Ask\_hydroxyl\_S.  
 DR InterPro; IPR00742; EGF-2.  
 DR InterPro; IPR001881; EGF-2\_Ca.  
 DR InterPro; IPR01438; EGF-III.  
 DR InterPro; IPR006309; EGF-like.  
 DR InterPro; IPR02049; Laminin\_EGF.  
 DR InterPro; IPR008297; Notch.  
 DR Inter-Pro; IPR00800; Notch\_dom.  
 DR Pfam; PF00003; ank; 6.  
 DR Pfam; PF00008; EGF; 35.  
 DR Pfam; PF00066; notch; 3.  
 DR Pfam; PF00022; Notch; 1.  
 DR PRINTS; PR0010; EGF\_BLOOD.  
 DR PRINTS; PR0152; NOTCH.  
 DR SMART; SM00248; ANK; 6.  
 DR SMART; SM00179; EGF; 23.  
 DR SMART; SM0004; NL; 3.  
 DR PROSITE; PS50297; ANK REP REGION; 1.  
 DR PROSITE; PS00088; ANK REPEAT; 4.  
 DR PROSITE; PS00110; ASX HYDROXYL; 20.  
 DR PROSITE; PS00022; EGF-1; 34.  
 DR PROSITE; PS00186; EGF-2; 26.  
 DR PROSITE; PS50026; EGF-3; 36.  
 DR PROSITE; PS0117; EGF-CA; 18.  
 DR Receptor; Transcription regulation; Activator; Differentiation;  
 DR Developmental protein; Repeat; ANK repeat; EGF-like domain;  
 DR Transmembrane; Glycoprotein; Signal; Phosphorylation.  
 FT SIGNAL  
 FT CHAIN 1 19 2556  
 FT CHAIN 172 2556  
 CC NOTCH EXTRACELLULAR TRUNCATION (By similarity).  
 CC NOTCH INTRACELLULAR DOMAIN (By

			Query Match	Score	DB	Length	Best Local Similarity	Pred.	No.	1..2;	15;	Mismatches	Indels	Gaps
FT	DOMAIN	19	1736	SIMILARITY).										
FT	TRANSMEM	1737	1757	EXTRACELLULAR (POTENTIAL).										
FT	DOMAIN	20	2556	POTENTIALIC (POTENTIAL).										
FT	DOMAIN	59	58	CYTOSLIC (POTENTIAL).										
FT	DOMAIN	99	99	EGF-LIKE 1.										
FT	DOMAIN	102	139	EGF-LIKE 2.										
FT	DOMAIN	140	176	EGF-LIKE 3.										
FT	DOMAIN	178	216	EGF-LIKE 4.										
FT	DOMAIN	218	255	EGF-LIKE 5. CALCIUM-BINDING (POTENTIAL).										
FT	DOMAIN	257	293	EGF-LIKE 6. CALCIUM-BINDING (POTENTIAL).										
FT	DOMAIN	295	333	EGF-LIKE 7. CALCIUM-BINDING (POTENTIAL).										
FT	DOMAIN	335	371	EGF-LIKE 8. CALCIUM-BINDING (POTENTIAL).										
FT	DOMAIN	372	410	EGF-LIKE 9. CALCIUM-BINDING (POTENTIAL).										
FT	DOMAIN	412	450	EGF-LIKE 10. CALCIUM-BINDING (POTENTIAL).										
FT	DOMAIN	452	488	EGF-LIKE 11. CALCIUM-BINDING (POTENTIAL).										
FT	DOMAIN	490	526	EGF-LIKE 12. CALCIUM-BINDING (POTENTIAL).										
FT	DOMAIN	528	564	EGF-LIKE 13. CALCIUM-BINDING (POTENTIAL).										
FT	DOMAIN	566	601	EGF-LIKE 14. CALCIUM-BINDING (POTENTIAL).										
FT	DOMAIN	603	639	EGF-LIKE 15. CALCIUM-BINDING (POTENTIAL).										
FT	DOMAIN	641	676	EGF-LIKE 16. CALCIUM-BINDING (POTENTIAL).										
FT	DOMAIN	678	714	EGF-LIKE 17. CALCIUM-BINDING (POTENTIAL).										
FT	DOMAIN	716	751	EGF-LIKE 18. CALCIUM-BINDING (POTENTIAL).										
FT	DOMAIN	753	789	EGF-LIKE 19. CALCIUM-BINDING (POTENTIAL).										
FT	DOMAIN	791	827	EGF-LIKE 20. CALCIUM-BINDING (POTENTIAL).										
FT	DOMAIN	829	868	EGF-LIKE 21. CALCIUM-BINDING (POTENTIAL).										
FT	DOMAIN	870	906	EGF-LIKE 22. CALCIUM-BINDING (POTENTIAL).										
FT	DOMAIN	908	944	EGF-LIKE 23. CALCIUM-BINDING (POTENTIAL).										
FT	DOMAIN	946	982	EGF-LIKE 24. CALCIUM-BINDING (POTENTIAL).										
FT	DOMAIN	984	1020	EGF-LIKE 25. CALCIUM-BINDING (POTENTIAL).										
FT	DOMAIN	1022	1058	EGF-LIKE 26. CALCIUM-BINDING (POTENTIAL).										
FT	DOMAIN	1060	1096	EGF-LIKE 27. CALCIUM-BINDING (POTENTIAL).										
FT	DOMAIN	1098	1144	EGF-LIKE 28. CALCIUM-BINDING (POTENTIAL).										
FT	DOMAIN	1146	1182	EGF-LIKE 29. CALCIUM-BINDING (POTENTIAL).										
FT	DOMAIN	1184	1220	EGF-LIKE 30. CALCIUM-BINDING (POTENTIAL).										
FT	DOMAIN	1222	1266	EGF-LIKE 31. CALCIUM-BINDING (POTENTIAL).										
FT	DOMAIN	1268	1306	EGF-LIKE 32. CALCIUM-BINDING (POTENTIAL).										
FT	DOMAIN	1308	1347	EGF-LIKE 33. CALCIUM-BINDING (POTENTIAL).										
FT	DOMAIN	1349	1385	EGF-LIKE 34. CALCIUM-BINDING (POTENTIAL).										
FT	DOMAIN	1388	1427	EGF-LIKE 35. CALCIUM-BINDING (POTENTIAL).										
PT	REPEAT	1446	1481	LIN/NOTCH 1. CALCIUM-BINDING (POTENTIAL).										
PT	REPEAT	1482	1523	LIN/NOTCH 2. CALCIUM-BINDING (POTENTIAL).										
PT	REPEAT	1524	1563	LIN/NOTCH 3. CALCIUM-BINDING (POTENTIAL).										
PT	REPEAT	1567	1597	ANK 1.										
PT	REPEAT	1597	1957	ANK 2.										
PT	REPEAT	1957	1991	ANK 3.										
PT	REPEAT	1995	2024	ANK 4.										
PT	REPEAT	2028	2057	ANK 5.										
PT	REPEAT	2050	2090	POLY-VAL.										
PT	REPEAT	1576	1579	POLY-ARG.										
PT	REPEAT	1662	1685	POLY-PRO.										
PT	REPEAT	1729	1732	POLY-ALA.										
PT	DOMAIN	1741	1744	POLY-GLU.										
PT	DOMAIN	1905	1905	POLY-GLY.										
PT	DOMAIN	2260	2263	POLY-GLN.										
PT	DOMAIN	2404	2407	POLY-PRO.										
PT	DOMAIN	2411	2418	POLY-SER.										
PT	DOMAIN	2522	2525	CLEAVAGE (BY FURIN-LIKE PROTEASE) (BY SIMILARITY).										
PT	SITE	1665	1666	CLEAVAGE (BY FURIN-LIKE PROTEASE) (BY SIMILARITY).										
FT	DISULFID	24	37	BY SIMILARITY.										
FT	DISULFID	31	46	BY SIMILARITY.										
FT	DISULFID	48	57	BY SIMILARITY.										
FT	DISULFID	63	74	BY SIMILARITY.										
FT	DISULFID	87	87	BY SIMILARITY.										
FT	DISULFID	89	98	BY SIMILARITY.										
FT	DISULFID	106	117	BY SIMILARITY.										
FT	DISULFID	111	127	BY SIMILARITY.										
FT	DISULFID	129	138	BY SIMILARITY.										
FT	DISULFID	144	155	BY SIMILARITY.										
FT	DISULFID	149	164	BY SIMILARITY.										
FT	DISULFID	166	175	BY SIMILARITY.										
FT	DISULFID	182	195	BY SIMILARITY.										

Query Match 3.0%; Score 145.5; DB 1; Length 2556;  
 Best Local Similarity 26.1%; Pred. No. 1..2;保守性 15; Mismatches 65; Indels 67; Gaps 11;  
 Matches 52; Conservative 15; Indels 208; Gaps 31;  
 Matches 127; Conservative 90; Mismatches 202; Indels 208; Gaps 31;  
 DR QY 435 VPNCAGCPGSWIKDGYDKACMNSACTWDGGDCSLSLNFDPKNCTQSLWCKYFSDGHC 1506  
 Db QY 1452 LPEQEDAGNK---WOSLCINHAGWGDGGS 533  
 DR QY 482 ---GGTGSIGVCPWPGGGINSVSCN---OGCANSWLADEKCFQACNTLSCERDAG 533  
 Db QY 1507 DSQCNAGCIFDGDCRAGEG---QNPPLYDQYCHDF---SDGHCDQGCNSAECENDGL 1560  
 DR QY 534 DCQDHFIE----LYVLLIP----NQHYLIPKGECU-----P 564  
 Db QY 1561 DCA-EHVPLAAGTLVVVLMPEQLRNSFHFLRLSRVLTNTVKRDAGQONIFP 1619  
 DR QY 565 YFSFAE----VAKRGVFG 578  
 Db QY 1620 YYGRBEEELRKHPITKRAANG 1638

RESULT 15

RN RN

GIDA\_BUCCP ID\_GIDA\_BUCCP STANDARD; PRT; 626 AA.

RX RX

AC P59485; DT 10-OCT-2003 (Rel. 42, Created)  
 RA 10-OCT-2003 (Rel. 42, Last sequence update)  
 RA 10-OCT-2003 (Rel. 42, Last annotation update)

DT DB Glucose inhibited division protein A.  
 DR GIDA OR BAF01.

RA Van Ham R.C.H.J., Kamerbeek J., Palacios C., Rausell C., Abascal F., Bastolla U., Fernandez J.M., Jimenez L., Postigo M., Silva F.J., Tamames J., Viguera E., Latorre A., Valencia A., Moran F., Moya A.; "Reductive genome evolution in Buchnera aphidicola"; Proc. Natl. Acad. Sci. U.S.A. 100:581-586(2003).

CC CC -!- FUNCTION: Not known.  
 CC CC -!- SIMILARITY: Belongs to the gida family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce> or send an email to license@isb-sib.ch).

CC DR EMBL; ABO14016; AA026745.1; -. DR HAMAP; MF\_00129; ; 1.  
 DR InterPro; IPR00218; GIDA.  
 DR InterPro; IPR004416; GIDA\_sub.  
 DR Pfam; PF01134; GIDA; 1.  
 DR TIGRFAMS; TIGR00136; gida; 1.  
 DR PROSITE; PS01280; GIDA; 1.  
 DR PROSITE; PS01281; GIDA; 2.  
 KW Complete proteome:  
 SEQUENCE 626 AA; 70393 MW; 9019A0DAE87C274 CRC64;

Query Match 3.0%; Score 145; DB 1; Length 626;  
 Best Local Similarity 20.3%; Pred. No. 0..16;保守性 90; Mismatches 202; Indels 208; Gaps 31;  
 Matches 127; Conservative 90; Mismatches 202; Indels 208; Gaps 31;  
 QY 212 LTDTKEVSVLQMLDPLSGRPPTFKEPNQTKPENLSSKVQKQLYSASVALKL 271  
 DR 153 LTDTKEVSVLQMLDPLSGRPPTFKEPNQTKPENLSSKVQKQLYSASVALKL 199

Search completed: July 26, 2004, 11:09:43  
Job time : 19.4568 secs

Copyright GenCore version 5.1.6  
(c) 1993 - 2004 Compugen Ltd.

Om protein - protein search, using sw model

Run on: July 26, 2004, 11:05:03 ; search time 60.6391 Seconds  
(without alignments)

4628.666 Million cell updates/sec

Title: US-10-657-280-1

Perfect score: 4907

Sequence: 1 MLLKLQRLQTYTCILSHRYGL.....SLKIQQAYFTDSKNTGQLK 928

Scoring table: BLASTM52

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 31518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTRMBL\_25:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phage:\*

10: sp\_plant:\*

11: sp\_rabbit:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_rvirus:\*

16: sp\_bacteriap:\*

17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

### SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	4473	91.2	847 4 Q86N13	Q96n13 homo sapien
2	3314	67.5	950 4 Q8UL2	Q9UL2 homo sapien
3	837	17.5	490 4 Q8TQ2	Q8TQ2 homo sapien
4	734.5	15.0	384 11 Q8D1340	Q8D1340 mus musculus
5	488	9.9	666 5 Q9V53	Q8X51 drosophila
6	360	7.3	71 4 Q8BWA5	Q9V53 drosophila
7	329	6.7	602 16 Q9L12	Q9BWA5 homo sapien
8	324.5	6.6	942 16 Q9L14	Q9L12 streptomyce
10	322.5	6.6	586 16 Q69853	Q9L14 streptomyce
11	320.5	6.5	486 2 Q7x4S1	Q69853 streptomyce
12	314	6.4	541 16 Q69851	Q7x4S1 neisseria m
13	287	5.8	238 2 Q8EVX1	Q69851 streptomyce
14	5.4	5.8	238 2 Q8GP3	Q8EVX1 streptococcus
15	266	5.4	545 2 Q8AD00	Q8GP3 streptococcus
16	5.4	5.4	545 2 Q83u59	Q8AD00 neisseria m

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

### ALIGNMENTS

RESULT 1

ID	Q96N13	PRELIMINARY;	PRT;	847 AA.
AC	Q96N13;			
DT	01-DEC-2001	(TREMBLrel. 19, Created)		
DT	01-DEC-2001	(TREMBLrel. 19, Last sequence update)		
DT	01-OCT-2003	(TREMBLrel. 25, Last annotation update)		
DE		Hypothetical Protein FLJ3575 (Fragment).		
OS		Human		
OC		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
OX				
NCBITaxID		3606;		
[1]				
RP		SEQUENCE FROM N.A.		
RA		Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,		
RA		Furuya T., Takahashi M., Kikkawa Y., Omura Y., Abe K., Kamihara K.,		
RA		Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,		
RA		Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isomoto Y.,		
RA		Kawai-Hig Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,		
RA		Matsuoka K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,		
RA		Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyan A., Kawakami B.,		
RA		Suzuki I., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.,		
RA		"NIHO" human cDNA sequencing project";		
RT		Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.		
RL		Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.		
DR		EMBL; AR056137; BAB71102.1; -.		
DR		GO: GO:0016020; C:membrane; IEA.		
DR		GO: GO:0030154; P:cell differentiation; IEA.		
DR		InterPro; IPR00800; Notch_dom.		
DR		PFAM; PF00066; notch; 1.		
DR		SMART; SM00004; NLU; 2.		
DR		Hypothetical protein.		
FT		NON_TER; 847		
SEQUENCE		847 AA; 95165 MW; 0A15015824733C5 CRC64;		
Query Match		91.2%; Score 4473; DB 4; Length 847;		
Best Local Similarity		99.8%; Pred. No. 7; ge-289;		
Matches		845; Conservative 1; Mismatches 1; Indexes 0; Gaps 0;		
QY		1 MLLKLQRLQTYTCILSHRYGLLYVCFGLGVVTIVSAFOFGEVLEWSRDQHVLFDSDYRDN1 60		

	RN [1]
DB	SEQUENCE FROM N.A.
QY	RP TISSUE:Brain;
DB	Medline:20039619; PubMed=10574462;
QY	RA Nagase T., Ishikawa K., Kikuno R., Hirose M., Nomura N., Ohara O.;
DB	"Prediction of the coding sequences of unidentified human genes. XV. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro.";
QY	RT DNA Res. 6:337-355(1999);
DB	DR EMBL: AB033034; BAA85222; -.
QY	DR GO; GO:0016020; C:membrane; IEA.
DB	DR GO; GO:000509; F:calcium ion binding; IEA.
QY	DR GO; GO:0031054; P:cell differentiation; IEA.
DB	DR InterPro: IPR000948; Ef-hand.
QY	DR InterPro: IPR000900; Notch_dom.
DB	DR Pfam: PF00056; ehand; 1.
QY	DR SMART; SM00004; NL: 2.
DB	DR PROSITE; PS00018; EF_HAND; 1.
QY	KW Hypothetical protein.
DB	FT 1
QY	TER 950 AA; 108785 MW; F653BDFAGEACT503 CRC64;
DB	SEQID NON_TER 1
QY	Query Match 67.5%; Score 3314; DB 4; Length 950;
DB	Best Local Similarity 99.8%; Pred. No. 1_2e-211; Mismatches 0; Indels 0; Gaps 0;
DB	Matches 621; Conservative 0; MisMatches 1; Indels 0; Gaps 0;
QY	307 SALQSOKODEDSASRFDENEELRSRSHIRIEGLSQKFYLNDVUMFCKDVMDDFVSHSKG 420
QY	421 YSHSKGQKVLTMPVPIAGEGPGSMWKGDYCDKACNNASCDWGDCDSGNSGGSRYIAG 480
DB	421 YSHSKGQKVLTMPVPIAGEGPGSMWKGDYCDKACNNASCDWGDCDSGNSGGSRYIAG 480
QY	367 IWTQDVERNLSHITPSSPAESHLIRIEGLSQKFYLNDVUMFCKDVMDDFVSHSKG 426
DB	481 GGTGSIISUGHMOPGGINSYSYCNOCANSWLAKEFDQAGNUISGEDAGDGQDHF 540
DB	481 GGTGSIISUGHMOPGGINSYSYCNOCANSWLAKEFDQAGNUISGEDAGDGQDHF 540
QY	541 HELKVKILLPNQHYYIPRGECLPYFFAEVAKRGVEGAYSDNPTRHASTANKKTHIL 600
DB	541 HELKVKILLPNQHYYIPRGECLPYFFAEVAKRGVEGAYSDNPTRHASTANKKTHIL 600
QY	601 IMSGMNMATIHNLFONTNDEBFKQIQTVEVDTREGPKNSTAQKGYENLVSPTTL 660
DB	601 IMSGMNMATIHNLFONTNDEBFKQIQTVEVDTREGPKNSTAQKGYENLVSPTTL 660
QY	651 EASILFEDLPKERCPKPKRHDUNSTRRAQEVRKIPVNLSLPPKQQLSNTLDQLEH 720
DB	651 EASILFEDLPKERCPKPKRHDUNSTRRAQEVRKIPVNLSLPPKQQLSNTLDQLEH 720
QY	721 GDTIKGNLSKALLSFLMSQHAKKNQAITDETDNSVAPQEKQHKSILPNSLG 780
DB	721 GDTIKGNLSKALLSFLMSQHAKKNQAITDETDNSVAPQEKQHKSILPNSLG 780
DB	301 NATTHENLTFONTNDEBFKQIQTVEVDTREGPKNSTAQKGYENLVSPTILP 666
QY	607 NATTHENLTFONTNDEBFKQIQTVEVDTREGPKNSTAQKGYENLVSPTILP 666
DB	241 ILLPNOTHYIPKGECLPYFSFAEVAKRGVEGAYSDNPTRHASTANKKTHILISMGM 606
DB	241 ILLPNOTHYIPKGECLPYFSFAEVAKRGVEGAYSDNPTRHASTANKKTHILISMGM 606
QY	667 EDTPKEKPFPKPKHDMNSTRRAQEVRKIPVNLSLPPKQQLSNTLDQLEHGTILK 726
DB	361 EDTPKEKPFPKPKHDMNSTRRAQEVRKIPVNLSLPPKQQLSNTLDQLEHGTILK 420
QY	727 GYNLSKALLSFLMSQHAKKNQAITDETDNSVAPQEKQHKSILPNSLG 786
DB	421 GYNLSKALLSFLMSQHAKKNQAITDETDNSVAPQEKQHKSILPNSLG 786
QY	787 RLTPEAVSVKUNGHDQGONPPDLETTARFRVETHQKTIQGANVTEKPSLIVPESOM 846
DB	481 RLTPEAVSVKUNGHDQGONPPDLETTARFRVETHQKTIQGANVTEKPSLIVPESOM 540
QY	847 TKEKKTGKEKENSMEENAENHGTVTBLGRKQYTDYLGFFPWEKKYFQDILDE 906
DB	541 TKEKKTGKEKENSMEENAENHGTVTBLGRKQYTDYLGFFPWEKKYFQDILDE 600
QY	907 BESEKTOQAYFTUSKGRLK 928
DB	601 BESEKTOQAYFTUSKGRLK 622
RESULT 2	RESULT 3
Q9ULL2 PRELIMINARY; PRT; 950 AA.	
ID Q9ULL2; DT 01-MAY-2000 (TREMBIREL 13, Created)	
DT 01-OCT-2002 (TREMBIREL 22, Last sequence update)	
DT 01-OCT-2003 (TREMBIREL 25, Last annotation update)	
DE Hypothetical protein KIAA1208 (Fragment).	
GN KIAA1208.	
OS Homo sapiens (Human).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
NCBI_TaxID=9606;	

Q86TQ2		PRELIMINARY;		PRT;	490 AA.
ID	Q86TQ2;				
AC	Q86TQ2;				
DT	01-JUN-2003	(TREMBrel. 24; Created)			
DT	01-JUN-2003	(TREMBrel. 24; Last sequence update)			
DE	similar to v-maf musculoaponeurotic fibrosarcoma oncogene family, DE protein B (Avian) (Fragment).				
OS	Homo sapiens (Human).				
OC	Eukaryota; Mettaoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Strausberg R.; TISSUE=Liver;				
RL	Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; BC042615; AAH4265.1; -.				
DR	GO:000509; F:calcium ion binding; IEA.				
DR	InterPro; IPR002048; ZF-hand.				
DR	PIR; PP00056; efnand; 1.				
DR	PROSITE; PS00018; EF_HAND; 1.				
FT	NON_TER_1				
SQ	SEQUENCE 490 AA; 57412 MW; 37B19FE0D1259AD2 CRC64;				
Query Match 17.1%; Score 837; DB 4; Length 490; Best Local Similarity 100.0%; Pred. No. 2.7e-47; Matches 162; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	767 EKQHVKSIPLNSLGVSERLQRITPAVSKVNGHDQGQNPPDLETTARPRVETHQKIL 826				
Db	1 EKQHVKSIPLNSLGVSERLQRITPAVSKVNGHDQGQNPPDLETTARPRVETHQKIL 60				
QY	827 GGNVTKKEPKSLVPLSOMTKERKIGTGKEKENSMEENAENHTGGTEVLGRKQHYTD 886				
Db	61 GGNVTKKEPKSLVPLSOMTKERKIGTGKEKENSMEENAENHTGGTEVLGRKQHYTD 120				
QY	887 SYLGFLPWEKKYKFQDLIDDEESLKTQIAYFTDSKNTGRQLK 928				
Db	121 SYLGFLPWEKKYKFQDLIDDEESLKTQIAYFTDSKNTGRQLK 162				
RESULT 4					
061340	PRELIMINARY; ID: Q61340; PRT; 384 AA.				
AC	Q61340; 01340; 01-NOV-1996 (TREMBrel. 01, Created)				
DT	01-NOV-1996 (TREMBrel. 01; Last sequence update)				
DT	01-OCT-2003 (TREMBrel. 25; Last annotation update)				
DE	basic domain/leucine zipper transcription factor (Fragment).				
GN	MABF.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10990; RN				
RP	SEQUENCE FROM N.A.				
RA	STRIK=Berkeley;				
RA	Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,				
RA	Champe M., Chavez C., Dorsett V., Dresnack D., Farfan D., Frise E.,				
RA	George R., Gonzalez M., Guarin H., Krcmiller B., Li P., Liao G.,				
RA	Miranda A., Gonzales C.J., Nunno J., Pacelli J., Paragas V., Park S.,				
RA	Patel S., Phuchanawong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,				
RA	Cainiker S., Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AV089618; AA19036-1; -.				
DR	PI-Base; FBgn003392; CG8027.				
DR	GO; GO:0016020; C:membrane; IEA.				
DR	GO; GO:0030154; P:cell differentiation; IEA.				
DR	InterPro; IPR000800; Notch_dom.				
DR	PIR; PF00066; notch; 1.				
SQ	SEQUENCE 666 AA; 77745 MW; F6FDB6D1C1C39248 CRC64;				
Query Match 9.9%; Score 488; DB 5; Length 666; Best Local Similarity 54.2%; Pred. No. 7.1e-24; Matches 84; Conservative 30; Mismatches 41; Indels 0; Gaps 0;					
QY	315 DEDTISASRBDDEEYLRSRSIERAHPWTNNTIVTNGQPSWLNLDNPRTVTHQDF 374				
Db	86 DDKDPSRFDDKNELYRSLSRSLERKAWIRKIVTNGQPSWLDLSYRVTVPHENLA 145				
QY	375 RNLSHLPTSSPALESHIRIEGSIKQFYIINDDMFGKDWPDYFSHSKGQYLTWP 434				
Db	146 PDPDPLPTSSSAETFLHRIPKLUSRFLNDIFLGAPLYPEDLYTAEGVRVYQAM 205				
QY	435 VPNCAGLEGCGSWIKDGYCMKACNSACDQDGDCS 469				
Db	206 VPGCLDPCPTVGGACDHCNTDACQDGDCS 240				



OC Bacteria; Actinobacteria; Actinomycetidae; Actinomycetales;  
 OC Streptomyces; Streptomycetaceae; Streptomyces.  
 RN NCBI\_TaxID=1902;  
 RP [1] SEQUENCE FROM N.A.  
 STRAIN=A3 (2) / M145;  
 RX MEDLINE=21986410; PubMed=12000953;  
 RA Bentley S.D., Chatter K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,  
 RA Rabinowitzsch E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 coelicolor A3(2)." ;  
 RL Nature 417:141-147 (2002).  
 DR EMBL; AL939113; CAB73573.1; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 602 AA; 67501 MW; 720EAB86A31D943F CRC64;  
 Query Match 6.7%; Score 329; DB 16; Length 602;  
 Best Local Similarity 33.6%; Pred. No. 2.3e-13;  
 Matches 79; Conservative 44; Mismatches 82; Indels 30; Gaps 5;  
 QY 290 INQKELTISPAVYLWDLSAISKSQDKDISASRFDENBEYLRSLSIERTAPWVENTY 349  
 DB 297 VDGSD---PANIRSRAFFSDPRTYRERAAANARYLSDEURSLSLNLYAPWVNLY 352  
 QY 350 TNGQIOPSWNLNLPNVTIVTHDVERMLSHPTFSSPAISHEIHTREGLOKETYLNDV 409  
 DB 353 TADQTPDWNTHDPRLKVSHKBIFSPBPTSLPTFNHSIAESOLHHIDGLSBHFLYFNDV 412  
 QY 410 MEGKDVWDDFVSHSKQOKVLT---WPNV-----CAEGCFSWI 447  
 DB 413 MIGRETQFQFLPGLNGQQYYLSPALIIPGEPENSEDPPVAAAGKONRRLAERFGSTI 472  
 QY 448 -KGKQDKACNN--SAGDWDGDCGSNSGASRVIAGGGTSIGVGHPMQFGGG 498  
 DB 473 RKKMHPVHALHRGVLAEITDFADEHRRTAASRFRSAGDVTSSLHVVAFHG 527  
 RESULT 9  
 QLL14 PRELIMINARY; PRT; 942 AA.  
 ID QLL14  
 AC Q9L114;  
 DT 01-OCT-2000 (TREMBrel. 15, Created)  
 DT 01-OCT-2000 (TREMBrel. 15, Last sequence update)  
 DT 01-JUN-2003 (TREMBrel. 24, Last annotation update)  
 DR Putative transferase.  
 GN SC02592 OR SCC88\_03.C  
 OS Streptomyces coelicolor.  
 OC Bacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycineae; Streptomycetaceae; Streptomyces.  
 RN NCBI\_TaxID=1902;  
 RP [1] SEQUENCE FROM N.A.  
 STRAIN=A3 (2);  
 RA Saunders D., Harris D.;  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.  
 RP [2] SEQUENCE FROM N.A.  
 STRAIN=A3 (2);  
 RA Cerdano-Tarraga A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RA Cerdano-Tarraga A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RA Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.  
 RP [3] SEQUENCE FROM N.A.  
 RN STRAIN=A3 (2);  
 RX MEDLINE=9700351; PubMed=8843316;  
 RA Redenbach M., Kieser H.M., Denapaitre D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RN SEQUENCE FROM N.A.  
 STRAIN=A3 (2) / M145;  
 RX MEDLINE=21986410; PubMed=12000953;

RT "A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 RT Mol. Microbiol. 21:77-96 (1996).  
 RN [4] SEQUENCE FROM N.A.  
 STRAIN=A3 (2) / M145;  
 RX MEDLINE=21986410; PubMed=12000953;  
 RA Bentley S.D., Chatter K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,  
 RA Rabinowitzsch E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 coelicolor A3(2)." ;  
 RL Nature 417:141-147 (2002).  
 DR EMBL; AL939113; CAB75373.1; -.  
 DR GO; GO:0009058; P-biosynthesis; IEA.  
 DR InterPro; IPR001996; Glyco\_transl.  
 DR Pfam; PF00534; Glycos transferase 1; 1.  
 KW Transferase; Complete proteome.  
 SQ SEQUENCE 942 AA; 105063 MW; 108B737B98A97B80 CRC64;  
 Query Match 6.6%; Score 324.5; DB 16; Length 942;  
 Best Local Similarity 28.1%; Pred. No. 8.9e-12;  
 Matches 93; Conservative 57; Mismatches 122; Indels 59; Gaps 11;  
 QY 175 NPSTNVSVVFDPSKDKVDAHSGKLKGNSRQTMWRGWTIDKVEPGVLMQ----DLA 228  
 DB 490 NPRLRLLAVDAEGR-VRKALAGAYE---KAVYAEILKPRTHAPGTLASERLEAVGEVA 545  
 QY 229 FLSGFPFPKETNQIK-----TKLPENLSSK-----VKKLQIYSE 263  
 DB 546 GLRFRPRVVTISTRTRGPGAYGDIEFWRQVDEEGDGQFVAPLRSAVGPKLPSLTD 605  
 QY 264 ASVALKLNKNPKOELNKQTK---NMTEIDKELTISPVLLN-----LSAT 309  
 DB 606 ARRIV---KDRYKPLTPIRKLVSLTIPFDVYVWDDSPRWRARRAAGGLE 661  
 QY 310 SQSKQDIDASIFPDNEELR-SIRSTERHAPWVNRIPIVINGQIWSWLNLPNPRVIVT 369  
 DB 662 AEEGDE---AAKPRNRDLELRSLRSLAMFAWIKRKYLVLQDTPWLNTHEGIIVVS 718  
 QY 370 HQDVFRRNISHLPPRSSPAISHEIHTREGLOKETYLNDVWGGKDWPPDVFSHKSQKV 429  
 DB 719 HRDIFTDQCLPTNSIESOLHHIQNSCQYLNNDVFRGPVGAQRFLPNGASRF 778  
 QY 430 YLT-MPVPRKAEGGSGWIKDGKDACKNSA 460  
 DB 779 F--WSPTTVPGEP-TEDEGYPAAKNRA 806  
 RESULT 10  
 QLL14 PRELIMINARY; PRT; 586 AA.  
 ID Q9L853  
 AC Q9853;  
 DT 01-AUG-1998 (TREMBrel. 07, Created)  
 DT 01-AUG-1998 (TREMBrel. 07, Last sequence update)  
 DT 01-JUN-2003 (TREMBrel. 24, Last annotation update)  
 DR Hypothetical protein SC06223.  
 GN SC0623 OR SC13.11.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycineae; Streptomycetaceae; Streptomyces.  
 RN [1] SEQUENCE FROM N.A.  
 STRAIN=A3 (2) / M145;  
 RX MEDLINE=21986410; PubMed=12000953;



Db	358 FMGKRPVAPHSFFI-PNGTARY--FPSSRNRIPOGAVAB--TDSPVDAACKNN 403
RESULT 13	<p>PRELIMINARY; PRT; 238 AA.</p> <p>QPEVKI; QPEVKI; 2001 (TREMBlrel. 16, Created)</p> <p>01-MAR-2001 (TREMBlrel. 16, Last sequence update)</p> <p>01-JUN-2003 (TREMBlrel. 24, Last annotation update)</p> <p>DE Patrative hexose transferase.</p> <p>GN CPSJ.</p> <p>OS Streptococcus salivarius.</p> <p>OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;</p> <p>OC Streptococcus.</p> <p>OX NCBI_TaxID:1304;</p> <p>RN [1]</p> <p>SEQUENCE FROM N.A.</p> <p>RX MEDLINE=20519245; PubMed=11065358;</p> <p>RA Almiron-Roig E., Mulholland F., Gasson M.J., Griffin A.M.; Microbiology 166:2793-2802(2000).</p> <p>RT "The complete cps gene cluster from Streptococcus thermophilus NCIB</p> <p>RT 2393 involved in the biosynthesis of a new exopolysaccharide.";</p> <p>RL DR EMBL; Y17900; CACB8360.1; -.</p> <p>DR GO: GO:0016740; F:transferase activity; IFA.</p> <p>RW Transferase.</p> <p>SQ SEQUENCE 238 AA; 29110 MW; A964250B01B78998 CRC64;</p>
QY	Query Match 5.8%; Score 287; DB 2; Length 238; Best Local Similarity 38.3%; Pred. No. 3.6e-11; Matches 59; Conservative 27; Mismatches 44; Indels 24; Gaps 3; ID 11 NDDEPFIREKPKVPHNRKID-----NDEP-NVHYRDGTENY 47
Db	<p>SLASIERHAPWWRNIFIVTINGQPSWLNLDNPRVTVTHQDVFERNISHLPLPTFSSPAIESH 391</p> <p>48 WFRMVERAHPWVNILITNGQPKWLNVNHFLKWKVRHEE-FIPKEYLPPFNASLEMN 106</p>
QY	IRHIEGLSOKFIFTNDYMFKGKOUWPPDFYSISK 425
Db	107 IHRIDGESENFLFNDMMLIGDVKYSDDFFVNNEK 140
RESULT 14	<p>PRELIMINARY; PRT; 238 AA.</p> <p>QGPDP3 QGPDP3 PRELIMINARY; PRT; 238 AA.</p> <p>DT 01-MAR-2003 (TREMBlrel. 23, Created)</p> <p>DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)</p> <p>DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)</p> <p>DE EBS5J.</p> <p>OS Streptococcus thermophilus.</p> <p>OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;</p> <p>OC Streptococcus.</p> <p>OX NCBI_TaxID:1308;</p> <p>RN [1]</p> <p>SEQUENCE FROM N.A.</p> <p>RALIU F., Erlich D.S., Renault P.; Diversity of cps operons in Streptococcus thermophilus.; Submitted (DBCC-2001) to the EMBL/GenBank/DBJ databases.</p> <p>SEQUENCE 238 AA; 29110 MW; A964250B01B78998 CRC64;</p>
QY	Query Match 5.8%; Score 287; DB 2; Length 238; Best Local Similarity 38.3%; Pred. No. 3.6e-11; Matches 59; Conservative 27; Mismatches 44; Indels 24; Gaps 3; ID 11 NDPERIREKQYVPHNRKID-----NDEP-NVHYRDGTENY 47
Db	<p>SLRSISTERHAPWWRNIFIVTINGQPSWLNLDNPRVTVTHQDVFERNISHLPLPTFSSPAIESH 391</p> <p>48 WFRMVERAHPWVNILITNGQPKWLNVNHFLKWKVRHEE-FIPKEYLPPFNASLEMN 106</p>
QY	IRHIEGLSOKFIFTNDYMFKGKOUWPPDFYSISK 425
Db	107 IHRIDGESENFLFNDMMLIGDVKYSDDFFVNNEK 140
RESULT 15	<p>PRELIMINARY; PRT; 545 AA.</p> <p>Q84D00 Q84D00 PRELIMINARY; PRT; 545 AA.</p> <p>ID Q84D00; AC 084D00; DT 01-JUN-2003 (TREMBlrel. 24, Created)</p> <p>DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)</p> <p>DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)</p> <p>DE SacB.</p> <p>GN SACB.</p> <p>OS Neisseria meningitidis.</p> <p>OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;</p> <p>OC Neisseriaceae; Neisseria.</p> <p>OX NCBI_TaxID:487;</p> <p>RN [1]</p> <p>SEQUENCE FROM N.A.</p> <p>RC STRAIN=M4775; Sacchi C.T., Whitney A.A., Mayer L.W., Mothershed E., Popovic T.;</p> <p>RT Submitted (FBB-2003) to the EMBL/GenBank/DBJ databases.</p> <p>RL EMBL; AY234202; AAC085300.1; -.</p> <p>SQ SEQUENCE 545 AA; 64134 MW; A3648A3C85AA946F CRC64;</p>
QY	Query Match 5.4%; Score 266; DB 2; Length 545; Best Local Similarity 27.8%; Pred. No. 3.1e-09; Matches 90; Conservative 53; Mismatches 129; Indels 52; Gaps 13; ID 97 LPNSNLTK--PACLICLESHKEDFLN-----KFLTISSENLKQLOQYKFNGQIKRPKS 145
Db	<p>LPANITLKVPSLXPSFHSAASDIFNVAKPKNPSTNVSVVFDSSTDQVEDAHGGILKG-NS 203</p> <p>97 LPNSNLTK--PACLICLESHKEDFLN-----KFLTISSENLKQLOQYKFNGQIKRPKS 145</p>
QY	204 RQTWGRGY-----LTDKVKPGLV-----LMQDIAFLSFRPPTKETWQIKT 245
Db	146 VNETWDLTSIAHVDMKUSTDRFLISSISQFWMLERCFEDKPI-----IPTANRYSR 200
QY	246 KLPENLSSKVKLUOLYSEASVALKLKNPKDFQELNKOTKKONTIDSKELTISPAYLMD 305
Db	201 KLWKH--SIKNWLFKEG-----IRNYESISSLPSYEEEDHNFIDLFLVFTWNSEDKNWO 251
QY	306 LSATISQSKOP--EDISASRFENBNELYRISLSTERHAPWWRNIFIVTINGQPSWLNLDN 362
Db	252 -ELVKYKDFENDATSTSFRSLERFLKALSWEMNGSFIRKIFIVSNCAAPPWDLUN 310
QY	363 PRVTVTHQDVFERNISHLPTFSSPAIESHIRTREGLSOKFIFTNDYMFKGKOUWPPDFYS 422
Db	311 PKIQWVYHEIMQ-SALPFTSSHAESLHHPGISMVFIYVNSDDFLLTFLINKDNFFY 369
QY	423 HSKGQKVYI-TWPVPN--CAEGCP 443
Db	370 SNGTAKLRLEAWGVNGCCTEGEP 393

This Page Blank (up to)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2004, 11:01:28 ; search time 30.2575 Seconds  
(without alignments) updates/sec  
3062.893 Million dels/sec

Title: US-10-657-280-2

Perfect score: 1731

Sequence: DTFADSLRVNKILNSKFGF.....RKIFPRRIKHEASPNRIRV 328

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

- 1: geneseq1980s:\*
- 2: geneseq1990s:\*
- 3: geneseq2000s:\*
- 4: geneseq2001s:\*
- 5: geneseq2002s:\*
- 6: geneseq2003as:\*
- 7: geneseq2004s:\*
- 8: geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

### SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	1731	100.0	328	4 ABR61377
2	1731	100.0	328	7 ADD27813
3	1731	100.0	328	7 ABW01489
4	1731	100.0	328	7 ABW01538
5	1731	100.0	1196	5 AAE25294
6	1731	100.0	1256	5 AAE25290
7	1698	98.1	328	4 ABW01536
8	1698	98.1	328	7 ADD27818
9	1698	98.1	328	7 ABW01492
10	1698	98.1	328	7 ABW01541
11	1487	85.9	1199	7 ADD27810
12	1487	85.9	1199	7 ABW01487
13	1487	85.9	1199	7 ABW01536
14	1480	65.9	1459	6 ABW01731
15	704	40.7	132	4 ABW01369
16	506	29.2	652	4 ABP5094
17	402	23.2	502	4 ABP61383
18	402	23.2	502	7 ADD27824
19	402	23.2	502	7 ABP01495
20	402	23.2	502	7 ABW01544
21	145	8.4	28	4 ABW61390
22	145	8.4	28	4 ABW6261
23	124	7.2	77	4 AAB88395
24	124	7.2	1312	5 ABP7794
25	124	7.2	1312	5 ABP53887

### ALIGNMENTS

RESULT 1

ID ABR61377

XX ABR61377 standard; protein; 328 AA.

XX AC

XX ABR61377;

DT 01-AUG-2003 (first entry)

XX DB Human

XX Human GlcNAc-phosphotransferase beta-subunit.

XX Human; N-Acetylglucosamine-1-phosphotransferase; nephrotropic; GlcNAc-phosphotransferase; phosphodiester alpha-N-GlcNAcase; N-acetylglucosamine-1-phosphodiester alpha-N-acetylglucosaminidase; enzyme replacement therapy; phosphorylated lysosomal hydrolase; lysosomal storage disease; enzyme; beta-subunit.

XX KW

XX Homo sapiens.

XX OS

XX PN US6537785-B1.

XX PD 25-MAR-2003.

XX PP 10-AUG-2000; 2000US-00636077.

XX PR 14-SEP-1999; 99US-0153831P.

XX PA (GENZ- ) GENZIME GLICOBILOGY RES INST INC.

XX PI Canfield WM;

XX DR WPI; 2001-290356/30.

XX N-RSDB; ACC81001.

XX PT Novel N-acetylglucosamine-1-phosphotransferase and N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase, useful for producing PT phosphorylated lysosomal hydrolase for treating lysosomal storage diseases.

XX PT Disclosure; Page 28-29; 62pp; English.

XX CC The invention relates to a novel isolated human N-acetylglucosamine-1-phosphotransferase (GlcNAc-phosphotransferase) (I) and phosphodiester

CC alpha-GlcNAcase (N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase) (II). The protein of the invention has

CC nephrotropic activity, and may be useful in enzyme replacement therapy. A protein of the invention (I), (II) is useful for preparing a phosphorylated lysosomal hydrolase. The phosphorylated hydrolase comprising a terminal mannose-6-phosphate, is useful for treating a

Aaw79296 Neisseria

Add27834 GlcNAc-ph

ADP26059 Streptoco

Abp49330 Listeria

Abp06505 Novel hum

Adm40883 Human pol

Aar79314 IL-3 cont

Aar79331 IL-3 cont

Aay53212 Human int

Aay53195 Human int

Aae14006 Myelopoie

Aae1398 Myelopoie

Aag97762 Human int

Aag97779 Human Int

Abp59432 Drosophil

Abg19268 Novel hum

Abg5330 Human pro

Adc31323 Human nov

Abg97482 Human Nov

Aam79838 Human Pro

CC patient suffering from a lysosomal storage disease. The present sequence  
 CC is used in the exemplification of the invention  
 XX

Sequence 328 AA;

Query Match	100.0%	Score 1731; DB 4; Length 328;
Best Local Similarity	100.0%	Pred. No. 2.3e-160; Mismatches 0; Indels 0; Gaps 0;
Matches	328;	Conservative 0; Sequence 328 AA;

```

  1 DTFADSLRUNKLNSKGFTSRKPVAPMMPHMIDRIVQELDMFPEFDXTSFHKYRHS 60
  1 DTFADSLRUNKLNSKGFTSRKPVAPMMPHMIDRIVQELDMFPEFDXTSFHKYRHS 60
  61 EDMQAFASFYFYIYMSAVOPLNISQVDFEVTDQSGVTSREIRTLATRTHELPLS1QDLT 120
  61 EDMQAFASFYFYIYMSAVOPLNISQVDFEVTDQSGVTSREIRTLATRTHELPLS1QDLT 120
  121 GLEMMLINCNSKMPADITOLNNIPPTQSYPPNLPPVTKSLVNTCKPVTIKHAYKD 180
  121 GLEMMLINCNSKMPADITOLNNIPPTQSYPPNLPPVTKSLVNTCKPVTIKHAYKD 180
  QY 181 NKYRFEEINGEEETAFKMRTRNISHVWVGOLDDIRKNPKFVCLNDNDHNHKAQTVKAVL 240
  Db 181 NKYRFEEINGEEETAFKMRTRNISHVWVGOLDDIRKNPKFVCLNDNDHNHKAQTVKAVL 240
  QY 241 RDYVESMFPPIPSQELPPEYRPRFLHMELOEWRAVKDKEFWTHCVLATIMFTFSFF 300
  Db 241 RDYVESMFPPIPSQELPPEYRPRFLHMELOEWRAVKDKEFWTHCVLATIMFTFSFF 300
  QY 301 AEQLALKRKKIFRRRKKEASPNRIRV 328
  Db 301 AEQLALKRKKIFRRRKKEASPNRIRV 328

```

RESULT 2

Query Match	100.0%	Score 1731; DB 7; Length 328;
Best Local Similarity	100.0%	Pred. No. 2.3e-160; Mismatches 0; Indels 0; Gaps 0;
Matches	328;	Conservative 0; Sequence 328 AA;

```

  1 DTFADSLRUNKLNSKGFTSRKPVAPMMPHMIDRIVQELDMFPEFDXTSFHKYRHS 60
  1 DTFADSLRUNKLNSKGFTSRKPVAPMMPHMIDRIVQELDMFPEFDXTSFHKYRHS 60
  61 EDMQAFASFYFYIYMSAVOPLNISQVDFEVTDQSGVTSREIRTLATRTHELPLS1QDLT 120
  61 EDMQAFASFYFYIYMSAVOPLNISQVDFEVTDQSGVTSREIRTLATRTHELPLS1QDLT 120
  121 GLEMMLINCNSKMPADITOLNNIPPTQSYPPNLPPVTKSLVNTCKPVTIKHAYKD 180
  121 GLEMMLINCNSKMPADITOLNNIPPTQSYPPNLPPVTKSLVNTCKPVTIKHAYKD 180
  QY 181 NKYRFEEINGEEETAFKMRTRNISHVWVGOLDDIRKNPKFVCLNDNDHNHKAQTVKAVL 240
  Db 181 NKYRFEEINGEEETAFKMRTRNISHVWVGOLDDIRKNPKFVCLNDNDHNHKAQTVKAVL 240
  QY 241 RDYVESMFPPIPSQELPPEYRPRFLHMELOEWRAVKDKEFWTHCVLATIMFTFSFF 300
  Db 241 RDYVESMFPPIPSQELPPEYRPRFLHMELOEWRAVKDKEFWTHCVLATIMFTFSFF 300
  QY 301 AEQLALKRKKIFRRRKKEASPNRIRV 328
  Db 301 AEQLALKRKKIFRRRKKEASPNRIRV 328

```

RESULT 3

Query Match	100.0%	Score 1731; DB 7; Length 328;
Best Local Similarity	100.0%	Pred. No. 2.3e-160; Mismatches 0; Indels 0; Gaps 0;
Matches	328;	Conservative 0; Sequence 328 AA;

```

  1 DTFADSLRUNKLNSKGFTSRKPVAPMMPHMIDRIVQELDMFPEFDXTSFHKYRHS 60
  1 DTFADSLRUNKLNSKGFTSRKPVAPMMPHMIDRIVQELDMFPEFDXTSFHKYRHS 60
  61 EDMQAFASFYFYIYMSAVOPLNISQVDFEVTDQSGVTSREIRTLATRTHELPLS1QDLT 120
  61 EDMQAFASFYFYIYMSAVOPLNISQVDFEVTDQSGVTSREIRTLATRTHELPLS1QDLT 120
  121 GLEMMLINCNSKMPADITOLNNIPPTQSYPPNLPPVTKSLVNTCKPVTIKHAYKD 180
  121 GLEMMLINCNSKMPADITOLNNIPPTQSYPPNLPPVTKSLVNTCKPVTIKHAYKD 180
  QY 181 NKYRFEEINGEEETAFKMRTRNISHVWVGOLDDIRKNPKFVCLNDNDHNHKAQTVKAVL 240
  Db 181 NKYRFEEINGEEETAFKMRTRNISHVWVGOLDDIRKNPKFVCLNDNDHNHKAQTVKAVL 240
  QY 241 RDYVESMFPPIPSQELPPEYRPRFLHMELOEWRAVKDKEFWTHCVLATIMFTFSFF 300
  Db 241 RDYVESMFPPIPSQELPPEYRPRFLHMELOEWRAVKDKEFWTHCVLATIMFTFSFF 300
  QY 301 AEQLALKRKKIFRRRKKEASPNRIRV 328
  Db 301 AEQLALKRKKIFRRRKKEASPNRIRV 328

```

Claim 8; SEQ ID NO 5; 55pp; English.

The invention relates to a method of phosphorylating a protein comprising contacting the protein with a soluble GlcNAc-phototransferase (UDP-N-

CC acetylglucosamine) and producing a phosphorylated protein. The method is useful for treating a patient suffering from a lysosomal storage disease e.g. Fabry's disease. The present sequence represents the amino acid sequence of the human GlcNAc-phototransferase beta subunit.

RESULT 4

Query Match	100.0%	Score 1731; DB 7; Length 328;
Best Local Similarity	100.0%	Pred. No. 2.3e-160; Mismatches 0; Indels 0; Gaps 0;
Matches	328;	Conservative 0; Sequence 328 AA;

```

  1 DTFADSLRUNKLNSKGFTSRKPVAPMMPHMIDRIVQELDMFPEFDXTSFHKYRHS 60
  1 DTFADSLRUNKLNSKGFTSRKPVAPMMPHMIDRIVQELDMFPEFDXTSFHKYRHS 60
  61 EDMQAFASFYFYIYMSAVOPLNISQVDFEVTDQSGVTSREIRTLATRTHELPLS1QDLT 120
  61 EDMQAFASFYFYIYMSAVOPLNISQVDFEVTDQSGVTSREIRTLATRTHELPLS1QDLT 120
  121 GLEMMLINCNSKMPADITOLNNIPPTQSYPPNLPPVTKSLVNTCKPVTIKHAYKD 180
  121 GLEMMLINCNSKMPADITOLNNIPPTQSYPPNLPPVTKSLVNTCKPVTIKHAYKD 180
  QY 181 NKYRFEEINGEEETAFKMRTRNISHVWVGOLDDIRKNPKFVCLNDNDHNHKAQTVKAVL 240
  Db 181 NKYRFEEINGEEETAFKMRTRNISHVWVGOLDDIRKNPKFVCLNDNDHNHKAQTVKAVL 240
  QY 241 RDYVESMFPPIPSQELPPEYRPRFLHMELOEWRAVKDKEFWTHCVLATIMFTFSFF 300
  Db 241 RDYVESMFPPIPSQELPPEYRPRFLHMELOEWRAVKDKEFWTHCVLATIMFTFSFF 300
  QY 301 AEQLALKRKKIFRRRKKEASPNRIRV 328
  Db 301 AEQLALKRKKIFRRRKKEASPNRIRV 328

```

Claim 10; Page 18-19; 46pp; English.

XX

The invention relates to a method for producing a high mannose glycoprotein. The method comprises: introducing and expressing a polynucleotide encoding a glycoprotein into a mammalian cell; culturing the cell in the presence of a lectin to obtain a lectin resistant cell; isolating the cell; culturing the cell in the presence of deoxymannoimuramycin and kifunensine to inhibit glycosylation of the glycoprotein; and collecting the glycoprotein. The invention is useful in gene therapy. The method is useful for producing a high mannose lysosomal storage disease. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-phosphotransferase beta subunit precursor

CC

lysosomal storage disease. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-phosphotransferase beta subunit precursor

CC

lysosomal storage disease. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-phosphotransferase beta subunit precursor

CC

lysosomal storage disease. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-phosphotransferase beta subunit precursor

CC

lysosomal storage disease. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-phosphotransferase beta subunit precursor

CC

lysosomal storage disease. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-phosphotransferase beta subunit precursor

CC

lysosomal storage disease. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-phosphotransferase beta subunit precursor

CC

lysosomal storage disease. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-phosphotransferase beta subunit precursor

CC

lysosomal storage disease. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-phosphotransferase beta subunit precursor

CC

lysosomal storage disease. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-phosphotransferase beta subunit precursor

CC

lysosomal storage disease. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-phosphotransferase beta subunit precursor

CC

lysosomal storage disease. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-phosphotransferase beta subunit precursor

CC

lysosomal storage disease. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-phosphotransferase beta subunit precursor

CC

lysosomal storage disease. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-phosphotransferase beta subunit precursor

CC

lysosomal storage disease. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-phosphotransferase beta subunit precursor

CC

lysosomal storage disease. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-phosphotransferase beta subunit precursor

CC

lysosomal storage disease. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-phosphotransferase beta subunit precursor

CC

lysosomal storage disease. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-phosphotransferase beta subunit precursor

CC

lysosomal storage disease. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-phosphotransferase beta subunit precursor

CC

lysosomal storage disease. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-phosphotransferase beta subunit precursor

CC

lysosomal storage disease. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-phosphotransferase beta subunit precursor

CC

lysosomal storage disease. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-phosphotransferase beta subunit precursor

CC

lysosomal storage disease. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-phosphotransferase beta subunit precursor

CC

lysosomal storage disease. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-phosphotransferase beta subunit precursor

CC

lysosomal storage disease. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-phosphotransferase beta subunit precursor

CC

lysosomal storage disease. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-phosphotransferase beta subunit precursor

CC

lysosomal storage disease. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-phosphotransferase beta subunit precursor

CC

lysosomal storage disease. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-phosphotransferase beta subunit precursor

CC

lysosomal storage disease. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-phosphotransferase beta subunit precursor

Query Match 100.0%; Score 1731; DB 7; Length 328;  
Best Local Similarity 100.0%; Pred. No. 2.3e-160; Indels 0; Gaps 0;  
Matches 328; Conservative 0; Mismatches 0; PT  
Oy 1 DTFADSRLRVNKLNKNSKGFTSRKVPAHMPMIDRIMQELQDMPEEFDKTSFHKYRS 60  
Db 61 EDMQFAFSFYIYIMSAVPLNISQVEDDVDTQSGQLSDRIRTLATRIHELPLSQDIT 120  
Db 61 EDMQFAFSFYIYIMSAVPLNISQVEDDVDTQSGQLSDRIRTLATRIHELPLSQDIT 120  
Qy 121 GLEHMILNCSKMLPADTQLNNIPPTSYDPNLPPVTKSLVNTCKPVTDKHAKVD 180  
Db 121 GLEHMILNCSKMLPADTQLNNIPPTSYDPNLPPVTKSLVNTCKPVTDKHAKVD 180  
Qy 181 NKYRFEIIGEEEAFKMRTRNTHVSHVQGDIDRKRKNPKVECLNDNDNHKAQTVAVL 240  
Db 181 NKYRFEIIGEEEAFKMRTRNTHVSHVQGDIDRKRKNPKVECLNDNDNHKAQTVAVL 240  
Db 241 RDYFESMFPPIPSQFELPREYRNFLHMELOEWRAYRDKLKFWTHCVLATLIMFTFSFF 300  
Oy 301 AEQIALAKKSKIPRRRKHEASPNRIV 328  
Db 301 AEQIALAKKSKIPRRRKHEASPNRIV 328  
Db 301 AEQIALAKKSKIPRRRKHEASPNRIV 328

RESULT 4  
ABW01538  
ID ABW01538 standard; protein: 328 AA.  
XX  
AC ABW01538;  
XX  
DT 15-JAN-2004 (first entry)  
XX  
DE Human GlcNAc-phosphotransferase beta subunit precursor protein.  
XX  
KW Glycoprotein; lectin; lysosomal storage disease; gastrointestinal; N-acetylglucosamine-1-phosphotransferase; enzyme; gene therapy; human.  
OS Homo sapiens.  
XX  
US2003124653-A1.  
PD 03-JUL-2003.  
XX  
PF 21-DEC-2001; 2001US-00023890.  
XX  
PR 21-DEC-2001; 2001US-00023890.  
XX  
PA (NOVA-) NOVAZYME PHARM INC.  
PI Canfield WM;  
XX  
DR WPI; 2003-810985//76.

Query Match 100.0%; Score 1731; DB 7; Length 328;  
Best Local Similarity 100.0%; Pred. No. 2.3e-160; Indels 0; Gaps 0;  
Matches 328; Conservative 0; Mismatches 0; PT  
Oy 1 DTFADSRLRVNKLNKNSKGFTSRKVPAHMPMIDRIMQELQDMPEEFDKTSFHKYRS 60  
Db 61 EDMQFAFSFYIYIMSAVPLNISQVEDDVDTQSGQLSDRIRTLATRIHELPLSQDIT 120  
Db 61 EDMQFAFSFYIYIMSAVPLNISQVEDDVDTQSGQLSDRIRTLATRIHELPLSQDIT 120  
Qy 121 GLEHMILNCSKMLPADTQLNNIPPTSYDPNLPPVTKSLVNTCKPVTDKHAKVD 180  
Db 121 GLEHMILNCSKMLPADTQLNNIPPTSYDPNLPPVTKSLVNTCKPVTDKHAKVD 180  
Qy 181 NKYRFEIIGEEEAFKMRTRNTHVSHVQGDIDRKRKNPKVECLNDNDNHKAQTVAVL 240  
Db 181 NKYRFEIIGEEEAFKMRTRNTHVSHVQGDIDRKRKNPKVECLNDNDNHKAQTVAVL 240  
Db 241 RDYFESMFPPIPSQFELPREYRNFLHMELOEWRAYRDKLKFWTHCVLATLIMFTFSFF 300  
Oy 301 AEQIALAKKSKIPRRRKHEASPNRIV 328  
Db 301 AEQIALAKKSKIPRRRKHEASPNRIV 328  
Db 301 AEQIALAKKSKIPRRRKHEASPNRIV 328

RESULT 5  
AAE25294  
ID AAE25294 standard; protein; 1196 AA.  
XX  
AC AAE25294;  
XX  
DT 30-OCT-2002 (first entry)  
XX  
DE Human nucleic acid-associated protein (NAAP-13).  
XX  
KW Human; nucleic acid-associated protein; NAAP-13; neurological disorder; arteriosclerosis; cancer; cell proliferative disorder; atherosclerosis; lymphoma; epilepsy; Alzheimer's disease; developmental; anticonvulsant; autoimmune disorder; AIDS; allergy; anaemia; stroke; malaria; leishmania; gene therapy; nootropic; neuroprotective; cerebroprotective; viricide; immunosuppressive; protozoaide; antimicrobial.  
XX  
OS Homo sapiens.  
XX  
Key location/Qualifiers  
FT Peptide 1..34  
FT Label= Signal\_Peptide  
FT Domain 1..19  
FT /note= "Cytosolic domain"  
FT Domain 20..42  
FT /note= "Transmembrane domain"

FT Protein 35. .1196  
 FT /note= "Mature human NAAP-13"  
 FT Domain 43. .1152  
 FT /note= "Non-cytosolic domain"  
 FT Domain 1153. .1175  
 FT /note= "Transmembrane domain"  
 FT Domain 116. .1196  
 FT /note= "Cytosolic domain"  
 PN WO200250279-A2.  
 XX  
 PD 27-JUN-2002.  
 PF 19-DEC-2001; 2001WO-US050256.  
 XX  
 PR 21-DEC-2000; 2000US-025771P.  
 PR 05-JAN-2001; 2001US-026008P.  
 PR 16-JAN-2001; 2001US-0262302P.  
 PR 22-JAN-2001; 2001US-0263823P.  
 PR 02-FEB-2001; 2001US-026608P.  
 PR 29-OCT-2001; 2001US-0348442P.  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Baughin MR, Lu Y, Arvizu C, Ramkumar J, Yao MG, Policky JL;  
 PI Walia NK, Tribouley KM, Yue H, Batra S, Ding L, Lal PG;  
 PI Borowsky ML, Lu DAM, Gandhi AR, Griffin JA, Xu Y, Azimai Y;  
 PI Gietzen KJ, Tang YT, Warren BA, Mason PM, Burford N, Bafalia AJA;  
 PI Lee EA, Yang J, Gorrod AE, Emerling BM, Marquis JP, Lee SY;  
 PI Swarnakar A, Reddy R;  
 DR WPI; 2002-519887/55.  
 XX  
 PT Nucleic acid associated proteins and nucleic acids for diagnosing,  
 PT treating and preventing cell proliferative (e.g. cancers), neurological  
 PT (e.g. epilepsy or stroke) or autoimmune disorders (e.g. AIDS).  
 XX  
 PS Claim 68; Page 169-172; 193pp; English.  
 XX  
 CC The invention relates to nucleic acid-associated proteins (NAAP) and  
 CC nucleic acids. The nucleic acid and amino acid sequences are useful for  
 CC diagnosing, treating and preventing cell proliferative e.g.  
 CC arteriosclerosis, atherosclerosis, lymphoma or cancers), neurological  
 CC (e.g. epilepsy, Alzheimer's disease or stroke), developmental and  
 CC autoimmune disorders (e.g. AIDS, allergies, or anaemia) or infections  
 CC (e.g. malaria, or leishmania), as well as in assessing the effects of  
 CC exogenous compound on the expression of nucleic acid and amino acid  
 CC sequences of nucleic acid-associated proteins. The invention is useful in  
 CC gene therapy. The present sequence is human NAAP-13  
 XX  
 SQ Sequence 1196 AA;  
 Query Match 100.0%; Score 1731; DB 5; Length 1196;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-159;  
 Matches 328; Conservative 0; Mismatches 0; Index 0; Gaps 0;  
 QV 1 DTFAADSLRIVKLNKINSKEGFTSRKVAPAHMPNDRIMQELQDMFPEEFDKTSFKVRS 60  
 DV 869 DTFAADSLRIVKLNKINSKEGFTSRKVAPAHMPNDRIMQELQDMFPEEFDKTSFKVRS 928  
 QV 61 EDQMFARFYFYMMMSAQQLPLNISQFEDVPTDQSGLSDREIRTAATRHLPLSILQDT 120  
 DV 929 EDQMFARFYFYMMMSAQQLPLNISQFEDVPTDQSGLSDREIRTAATRHLPLSILQDT 988  
 QV 121 GLPHMLNCSKMLPADTQLNNIPPTOESYDPNLPPVTKSLVNTCKPKVDKPKVHAYOK 180  
 DV 989 GLPHMLNCSKMLPADTQLNNIPPTOESYDPNLPPVTKSLVNTCKPKVDKPKVHAYOK 1048  
 XX  
 PT Nucleic acid associated proteins and nucleic acids for diagnosing,  
 PT treating and preventing cell proliferative (e.g. cancers), neurological  
 PT (e.g. epilepsy or stroke) or autoimmune disorders (e.g. AIDS).  
 XX  
 PS Claim 64; Page 163-165; 193pp; English.  
 XX  
 CC The invention relates to nucleic acid-associated proteins (NAAP) and  
 CC nucleic acids. The nucleic acid and amino acid sequences are useful for  
 CC diagnosing, treating and preventing cell proliferative e.g.  
 CC arteriosclerosis, atherosclerosis, lymphoma or cancers), neurological  
 CC (e.g. epilepsy, Alzheimer's disease or stroke), developmental, and  
 CC autoimmune disorders (e.g. AIDS, allergies, or anaemia) or infections

QY 241 RDFFYSMFPPIPSPQELPREYRNFLPMELQENPRAYDGLKEWTHCYLATLNFITSFPE 300  
 Db 1109 RDFFYSMFPPIPSPQELPREYRNFLPMELQENPRAYDGLKEWTHCYLATLNFITSFPE 1168  
 QY 301 AEQLALKKIKPFRRRIHEASPRIRV 328  
 Db 1169 AEQLALKKIKPFRRRIHEASPRIRV 1196

RESULT 6  
 ID AAE25290  
 ID AAE55290 standard; protein; 1256 AA.  
 XX  
 AC AAE25290;  
 XX  
 DT 30-OCT-2002 (first entry)  
 XX  
 DE Human nucleic acid-associated protein (NAAP-9).  
 XX  
 PR Human; nucleic acid-associated protein; NAAP-9; neurological disorder;  
 PR arteriosclerosis; cancer; cell proliferative disorder; atherosclerosis;  
 PR lymphoma; epilepsy; Alzheimer's disease; developmental; anticonvulsant;  
 PR autoimmune disorder; AIDS; allergy; anaemia; stroke; malaria; leishmania;  
 PR gene therapy; nootropic; neuroprotective; cerebroprotective; viricide;  
 PR immunosuppressive; protozoocide; antimicrobial.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Domain 448. .469 /note= "Notch domain"  
 FT Domain 500. .536 /note= "Notch domain"  
 FT Domain 1018. .1030 /note= "NF-hand calcium-binding domain"  
 FT Domain  
 FT Domain  
 FT Domain  
 PN WO200250279-A2.  
 XX  
 PD 27-JUN-2002.  
 XX  
 PF 19-DEC-2001; 2001WO-US050256.  
 XX  
 PR 21-DEC-2000; 2000US-025774P.  
 PR 05-JAN-2001; 2001US-026008P.  
 PR 16-JAN-2001; 2001US-0262302P.  
 PR 23-JAN-2001; 2001US-026823P.  
 PR 02-FEB-2001; 2001US-026608P.  
 PR 29-OCT-2001; 2001US-0348442P.  
 PR  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Baughin MR, Lu Y, Arvizu C, Ramkumar J, Yao MG, Policky JL;  
 PI Walia NK, Tribouley KM, Yue H, Batra S, Ding L, Lal PG;  
 PI Borowsky ML, Lu DAM, Gandhi AR, Griffin JA, Xu Y, Azimai Y;  
 PI Gietzen KJ, Tang YT, Warren BA, Mason PM, Burford N, Bafalia AJA;  
 PI Lee EA, Yang J, Gorrod AE, Emerling BM, Marquis JP, Lee SY;  
 PI Swarnakar A, Reddy R;  
 DR WPI; 2002-519887/55.  
 XX  
 PT Nucleic acid associated proteins and nucleic acids for diagnosing,  
 PT treating and preventing cell proliferative (e.g. cancers), neurological  
 PT (e.g. epilepsy or stroke) or autoimmune disorders (e.g. AIDS).  
 XX  
 PS Claim 64; Page 163-165; 193pp; English.

(e.g. malaria, or leishmania), as well as in assessing the effects of exogenous compound on the expression of nucleic acid and amino acid sequences of nucleic acid-associated proteins. The invention is useful in gene therapy. The present sequence is human NAAP-9

Sequence 1256 AA;

Query Match 100.0%; Score 1731; DB 5; Length 1256;  
Best Local Similarity 100.0%; Pred. No. 1.6e-159; Indels 0; Gaps 0;  
Matches 328; Conservative 100.0%; Mismatches 0;

OY 1 DTPADSLRVNKILNSKGFTSRKVAPAHMIDRIVQMLQDMPEEFKTSFHKVRS 60  
Db 929 DTPADSLRVNKILNSKGFTSRKVAPAHMIDRIVQMLQDMPEEFKTSFHKVRS 988

OY 61 EDQGAFASFYIYLMSAVQPLNISQVFDVTQDQSGVLSDABRTLATHRHELPQLDT 120  
Db 989 EDQGAFASFYIYLMSAVQPLNISQVFDVTQDQSGVLSDABRTLATHRHELPQLDT 1048

Db 121 GLEMMLINCSKMLPADTQTLNNIPPTQSYDTPNLPPTVLSLVTCKPVDKIHAKYOK 180

OY 1049 GLEMMLINCSKMLPADTQTLNNIPPTQSYDTPNLPPTVLSLVTCKPVDKIHAKYOK 1108

OY 181 NKYRFEINGEEBETAFKMRITNTVSHVVGOLDDIRKNPRKFVCLNDIDNHKDQTKVAKL 240

Db 1109 NKYRFEINGEEBETAFKMRITNTVSHVVGOLDDIRKNPRKFVCLNDIDNHKDQTKVAKL 1168

OY 241 RDFFESMPPIPSQFELPPEYRNFLAMEHQEWRAVRDKLKFWMTCVALIMPTFSF 300

Db 1169 RDFFESMPPIPSQFELPPEYRNFLAMEHQEWRAVRDKLKFWMTCVALIMPTFSF 1228

OY 301 AEQITALAKRKIKFPRRKIHKEASPNRIRV 328

Db 1229 AEQITALAKRKIKFPRRKIHKEASPNRIRV 1256

RESULT 7

ABR61380 ABR61380 standard; protein; 328 AA.

AC ABR61380; DT 01-AUG-2003 (first entry)

XX Murine GlcNAc-phosphotransferase beta-subunit.

XX Mouse: N-acetylglucosamine-1-phosphotransferase; nephrotropic; GlcNAc-phosphotransferase; phosphodiester alpha-GlcNAcase; KW N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase; enzyme replacement therapy; phosphorylated lysosomal hydrolase; KW lysosomal storage disease; enzyme; beta-subunit. XX OS Mus musculus.

XX US6537785-B1. PN 25-MAR-2003.

XX PF 10-AUG-2000; 2000US-00636077. PR 14-SEP-1999; 99US-0153831P.

XX PA (GENZ-) GENZYME GLYCobiology RES INST INC.  
PT Canfield WM; XX WI; 2001-290356/30. DR N-PSDB; ACC81007.

XX Novel N-acetylglucosamine-1-phosphotransferase and N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase, useful for producing phosphorylated lysosomal hydrolase for treating lysosomal storage diseases.

XX Disclosure; Page 35-36; 62pp; English.  
XX The invention relates to a novel isolated human N-acetylglucosamine-1-phosphotransferase (GlcNAc-phosphotransferase) (I) and phosphodiester alpha-GlcNAcase (N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase) (III). The protein of the invention has nephropoietic activity, and may be useful in enzyme replacement therapy. A protein of the invention (I), (II) is useful for preparing a phosphorylated lysosomal hydrolase. The phosphorylated hydrolase comprising a terminal mannose-6-phosphate, is useful for treating a patient suffering from a lysosomal storage disease. The present sequence is used in the exemplification of the invention

XX Sequence 328 AA;

Query Match 98.1%; Score 1698; DB 4; Length 328;  
Best Local Similarity 97.6%; Pred. No. 3.9e-157; Indels 0; Gaps 0;  
Matches 320; Conservative 7; Mismatches 1;

OY 1 DTPADSLRVNKILNSKGFTSRKVAPAHMIDRIVQMLQDMPEEFKTSFHKVRS 60

Db 1 DT PADSLRVNKILNSKGFTSRKVAPAHMIDRIVQMLQDMPEEFKTSFHKVRS 60

OY 61 EDQGAFASFYIYLMSAVQPLNISQVFDVTQDQSGVLSDABRTLATHRHELPQLDT 120

Db 121 GLEMMLINCSKMLPADTQTLNNIPPTQSYDTPNLPPTVLSLVTCKPVDKIHAKYOK 180

OY 61 EDQGAFASFYIYLMSAVQPLNISQVFDVTQDQSGVLSDABRTLATHRHELPQLDT 120

Db 121 GLEMMLINCSKMLPADTQTLNNIPPTQSYDTPNLPPTVLSLVTCKPVDKIHAKYOK 180

OY 181 NKYRFEINGEEBETAFKMRITNTVSHVVGOLDDIRKNPRKFVCLNDIDNHKDQTKVAKL 240

Db 181 NKYRFEINGEEBETAFKMRITNTVSHVVGOLDDIRKNPRKFVCLNDIDNHKDQTKVAKL 240

OY 241 RDFFESMPPIPSQFELPPEYRNFLAMEHQEWRAVRDKLKFWMTCVALIMPTFSF 300

Db 301 AEQITALAKRKIKFPRRKIHKEASPNRIRV 328

Db 301 AEQITALAKRKIKFPRRKIHKEASPNRIRV 328

RESULT 8

ADD27818 ADD27818 standard; protein; 328 AA.

AC ADD27818; DT 15-JAN-2004 (first entry)

XX DE GlcNAc-phosphotransferase associated protein #2.

XX mouse; protein phosphorylation; soluble GlcNAc-phosphotransferase; ds.

XX UDP-N-acetylglucosamine; lysosomal storage disease; Fabry's disease; ds.

XX OS Mus musculus.

XX US2003119088-A1.

XX PD 26-JUN-2003.

XX PF 21-DEC-2001; 2001US-00023888.

XX PR 21-DEC-2001; 2001US-00023888.

XX PA (NOVA-) NOVAZYME PHARM INC.

XX PI Canfield W, Kudo M;

XX WPI; 2003-801323/75.

XX PT phosphorylating a protein for treating a patient suffering from a  
 XX lysosomal storage disease e.g. Fabry's disease by contacting the protein  
 PT with a soluble GlcNAc-phosphotransferase and producing a phosphorylated  
 PT protein.  
 XX  
 PS Disclosure; SEQ ID NO 10; 55pp; English.  
 XX  
 CC The invention relates to a method of phosphorylating a protein comprising  
 CC contacting the protein with a soluble GlcNAc-phosphotransferase (NDP-N-  
 acetyleglucosamine) and producing a phosphorylated protein. The method is  
 useful for treating a patient suffering from a lysosomal storage disease  
 e.g. Fabry's disease. The present sequence represents the amino acid  
 CC sequence of a GlcNAc phosphotransferase associated protein.  
 XX  
 SQ Sequence 328 AA;  

Query	Match	98.1%	Score	1698	DB	7	Length	328
QY	Best Local Similarity	97.6%	Pred.	No. 3	9e-157;			
	Matches	320;	Conservative			7;	Mismatches	1;
							Indels	0;
							Gaps	
Db	1	DTPADASRYVNKTILNSKRGFTSCKPVAIMPMTDRIVQBLQDMPPBFDKTSFKYTHS	60					
QY	1	EDNPQFATSYFYIYLMSAVOPLNISQVRFIEDTDSGVLSDREITLARTRIDPLSQDIT	122					
Db	1	DFPADSLRVNKLN SKRGFTSKVPAIMPMDRIVQBLQDMPPBFDKTSFKYTHS	60					
QY	61	EDNPQFATSYFYIYLMSAVOPLNISQVRFIEDTDSGVLSDREITLARTRIDPLSQDIT	122					
QY	61	EDNPQFATSYFYIYLMSAVOPLNISQVRFIEDTDSGVLSDREITLARTRIDPLSQDIT	122					
Db	121	GLEHMILNCSKMLPADTOLNNIPPTESYYPPNLPPVTKSVTNCVKPTVDKHKAYKD	18					
Db	121	GLEHMILNCSKMLPAT-TOLNNIPPTESYYPPNLPPVTKSVTNCVKPTVDKHKAYKD	18					
QY	181	NKYRFEINGEETAPKMRINSHVVGOLDIKRNPKVFVCLNDNHNKAQTVKAVL	241					
QY	181	NKYRFEINGEETAPKMRINSHVVGOLDIKRNPKVFVCLNDNHNKAQTVKAVL	241					
Db	241	RDFYSEMPPIPQSFELPREYRFLHMELOSWRAYDXKLKEWTHCYVLAUTMFTFSFF	300					
Db	241	RDFYSEMPPIPQSFELPREYRFLHMELOSWRAYDXKLKEWTHCYVLAUTMFTFSFF	300					
QY	301	ABQLAKRKTKPRRHKEASPNRIRV	328					
Db	301	AEQIALAKRKTKPRRHKEASPNRIRV	328					
	-	-	-					

RESULT 9  
 ABW01492  
 ID ABW01492 standard; protein: 328 AA.  
 XX  
 AC ABW01492;  
 XX  
 DT 15-JAN-2004 (first entry)  
 DE  
 Mouse protein #2 used to illustrate the method of the invention.  
 XX  
 KM Mannose glycoprotein; gene therapy; carbohydrate deficient cell;  
 KW lysosomal storage disease; gastrointestinal; mouse;  
 KW lectin resistant cell; deoxymannojirimycin; kifunensine;  
 KW glycosylation inhibition.  
 XX  
 OS Mus musculus.  
 XX  
 PN US2003124652-A1.  
 XX  
 PD 03-JUL-2003.  
 PP 21-DEC-2001; 2001US-00023889.  
 XX  
 PR (NOVA-) NOVAZYME PHARM INC.  
 PA  
 XX

PI Canfield WM;  
 XX DR WPI; 2003-810984/76.  
 PT Producing a high mannose glycoprotein for treating lysosomal storage  
 disease, comprises culturing the lectin resistant mammalian cell in the  
 presence of deoxymannojirimycin and kifunensine.  
 XX PS Disclosure; Page 25-26; 46pp; English.  
 XX SQ The invention relates to a method for producing a high mannose  
 glycoprotein. The method comprises: introducing and expressing a  
 polynucleotide encoding a glycoprotein into a mammalian cell; culturing  
 the cell in the presence of a lectin to obtain a lectin resistant cell;  
 isolating the cell; culturing the cell in the presence of  
 deoxymannojirimycin and kifunensine to inhibit glycosylation of the  
 glycoprotein; and collecting the glycoprotein. The invention is useful in  
 gene therapy. The method is useful for producing a high mannose  
 glycoprotein in a complex carbohydrate deficient cell for treating  
 lysosomal storage disease. The present sequence is mouse protein used to  
 illustrate the method of the invention  
 XX Sequence 328 AA;  
 Query Match 98.1%; Score 1698; DB 7; Length 328;  
 Best Local Similarity 96.6%; Pred. 3.9e-157; Matches 320; Conservative 7; Mismatches 1; Indels 0; gaps 0;  
 Matches 320; Conservative 7; Mismatches 1; Indels 0; gaps 0;  
 Oy 1 DTFAFDLSRTRYNKLNKNSKGRTSRSKPAIMPHMDRIVNOELDMFPERPDKTSFHKYRHs 60  
 Db 1 DTFASLRSVVKNLNSKGFTSKPAIMPHMDRIVNOELDMFPERPDKTSFHKYRHs 60  
 Qy 61 EDQFFAFPSFYIYLMSAVPLNTSQVFDVTDSQGVLSDRERTLATRIHPLSQTDLT 120  
 Db 61 EDMQFAPSFYIYLMSAVPLNTSQVFDVTDSQGVLSDRERTLATRIHPLSQTDLT 120  
 121 GLEMLINLNSKMLPADTQLNNIPPTQPSYDNPVTKSLVTKNCPVTDKIKAKD 180  
 121 GLEMLINLNSKMLPADTQLNNIPPTQPSYDNPVTKSLVTKNCPVTDKIKAKD 180  
 Qy 181 NCKRFPEIMGEETAAFKMTRNTSHVQGOLDDIRKRNPRKPVCLNDNDHDKDAQTWAVL 240  
 Db 181 NCKRFPEIMGEETAAFKMTRNTSHVQGOLDDIRKRNPRKPVCLNDNDHDKDAQTWAVL 240  
 241 RDFYESMPKIPSPOLPREYRFLHMELOWRAYDQLKLUWTHYWTQYLTIMPFTSFPP 300  
 241 RDFYESMPKIPSPOLPREYRFLHMELOWRAYDQLKLUWTHYWTQYLTIMPFTSFPP 300  
 Qy 301 AEQLALKTRKFRRTHKEASNRIRV 328  
 Db 301 AEQLALKTRKFRRTHKEASNRIRV 328  
 RESULT 10  
 ABW01541 ID ABW01541 standard; protein: 328 AA.  
 XX AC ABW01541;  
 XX ABW01541;  
 DT 15-JAN-2004 (first entry)  
 XX DE Mouse protein #2 used to illustrate the method of the invention.  
 KW Glycoprotein; lectin; lysosomal storage disease; gastrointestinal;  
 KW N-acetylglucosamine-1-phosphotransferase; gene therapy; mouse.  
 XX OS Mus musculus.  
 XX PN US2003124653-A1.  
 XX PD 03-JUL-2003.  
 XX  
 21-DEC-2001; 2001US-00023890.  
 PF

XX  
PR 21-DEC-2001; 2001US-00023890.  
PA XX  
(NOVA-) NOVAZYME PHARM INC.  
XX  
PI Canfield WM;  
XX  
WPI; 2003-810985/76.  
XX  
DR XX  
PT XX  
PT producing a glycoprotein with reduced complex carbohydrates by culturing the lectin resistant mammalian cell expressing the glycoprotein for treating lysosomal storage disease.  
XX  
PS Disclosure; Page 25-26; 46pp; English.  
XX  
CC The present invention provides a method of producing a glycoprotein having reduced complex carbohydrates by culturing the lectin resistant mammalian cell expressing the glycoprotein. The method is useful for producing a glycoprotein with reduced complex carbohydrates for treating lysosomal storage disease. The present invention is also useful in gene therapy. The present sequence is mouse protein used to illustrate the method of the invention  
CC Sequence 328 AA;  
XX  
Query Match 98.1%; Score 1698; DB 7; Length 328;  
Best Local Similarity 97.6%; Pred. No. 3.9e-157; Mismatches 7; Indels 0; Gaps 0; Matches 320; Conservative 0; MisMatches 1; Indels 0; Gaps 0;  
OY 1 DTFADSRLRVNKILNSKGFTSKRVPAMPHMDR-WMQETQDMPEEFDKTSFHKVRS 60  
Db 1 DTFADSRLRVNKILNSKGFTSKRVPAMPHMDR-WMQETQDMPEEFDKTSFHKVRS 60  
OY 61 EDMQFAFSYYIYIMSAVQPLNISQVFEDVDTQSGVLSDRRTLARTRHPLSQQDT 120  
Db 61 EDMQFAFSYYIYIMSAVQPLNISQVFEDVDTQSGVLSDRRTLARTRHPLSQQDT 120  
OY 121 GLEMHLINCSKOMPADTQLNNIPTPESYDPLNPVPTKSLVTCVKPVTDKIHAKYK 180  
Db 121 GLEMHLINCSKOMPADTQLNNIPTPESYDPLNPVPTKSLVTCVKPVTDKIHAKYK 180  
OY 181 NKYRFEEIGEEBAAFKMTRNTVSHVQGQDDTRKNPKFVCLNDNDHNDHDKAQTVKAVL 240  
Db 181 NKYRFEEIGEEBAAFKMTRNTVSHVQGQDDTRKNPKFVCLNDNDHNDHDKAQTVKAVL 240  
OY 241 RDYFESMPPIPQSQELPREYRNFLMHEQBWRAYDKLK 300  
Db 241 RDYFESMPPIPQSQELPREYRNFLMHEQBWRAYDKLK 300  
OY 301 AEQLIAKRKTKPFRRIKHEASPNRKV 328  
Db 301 AEQLIAKRKTKPFRRIKHEASPNRKV 328  
  
RESULT 11  
ADD27810 ID ADD27810 standard; protein; 1199 AA.  
AC XX  
AC ADD27810;  
XX  
DT 15-JAN-2004 (first entry)  
DE Soluble human GlcNAc-phosphotransferase.  
XX  
DE human; protein phosphorylation; soluble GlcNAc-phosphotransferase; UDP-N-acetylglucosamine; lysosomal storage disease; Fabry's disease; synthetic.  
OS Homo sapiens.  
OS  
XX PN US2003119088-A1.  
XX  
  
RESULT 12  
ABW01487 ID ABW01487 standard; protein; 1199 AA.  
AC XX  
AC ABW01487;  
XX  
DT 15-JAN-2004 (first entry)  
DE N-acetylglucosamine-1(GlcNAc)-phosphotransferase.  
XX  
DE Mannose glycoprotein; gene therapy; carbohydrate deficient cell; lysosomal storage disease; N-acetylglucosamine-1-phosphotransferase; gastrointestinal; enzyme; lectin resistant cell; deoxymannojirimycin; kifunensine; glycosylation inhibition.  
OS Unidentified.  
OS  
XX PN US2003124652-A1.

PD 03-JUL-2003.  
 XX  
 PR 21-DEC-2001; 2001US-00023899.  
 XX  
 PR 21-DEC-2001; 2001US-00023899.  
 XX  
 (NOVA-) NOVAZYME PHARM INC.  
 XX  
 PI Canfield WM;  
 XX  
 DR WPI; 2003-810984/76.  
 XX  
 DR N-PSDB; AAD62490.  
 XX  
 PT Producing a high mannose glycoprotein for treating lysosomal storage disease, comprising culturing the lectin resistant mammalian cell in the presence of deoxymannojirimycin and kifunensine.  
 XX  
 PS Claim 8; Page 10-13; 46pp; English.

CC The invention relates to a method for producing a high mannose glycoprotein. The method comprises: introducing and expressing a polynucleotide encoding a glycoprotein into a mammalian cell; culturing the cell in the presence of a lectin to obtain a lectin resistant cell; isolating the cell; culturing the cell in the presence of deoxymannojirimycin and kifunensine to inhibit glycosylation of the glycoprotein; and collecting the glycoprotein. The invention is useful in gene therapy. The method is useful for producing a high mannose glycoprotein in a complex carbohydrate deficient cell for treating lysosomal storage disease. The present sequence is N-acetylglucosamine-1 (GlcNAc)-phosphotransferase

XX Sequence 1199 AA;

Query Match 85.9%; Score 1487; DB 7; Length 1199;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-135;  
 Matches 281; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRIVNKLNKSGFTSRKVAPRHMDRTWQELDMFPEEFDKTSFHVRHS 60  
 Db 919 DTFADSLRIVNKLNKSGFTSRKVAPRHMDRTWQELDMFPEEFDKTSFHVRHS 978  
 QY 61 EDQFAFSFYIYMSAVOPLNTSQVFEVDTQSGVLSDRECFTRIHLPLS-QDLT 120  
 Db 979 EDQFAFSFYIYMSAVOPLNTSQVFEVDTQSGVLSDRECFTRIHLPLS-QDLT 1038  
 Db 121 GLEMMLINGSKMVPADTQLNNTIPPTQSYVPPNLPPTVKSLVTNCKPVTDKHKAYDK 180  
 QY 121 GLEMMLINGSKMVPADTQLNNTIPPTQSYVPPNLPPTVKSLVTNCKPVTDKHKAYDK 180  
 Db 1039 GLEMMLINGSKMVPADTQLNNTIPPTQSYVPPNLPPTVKSLVTNCKPVTDKHKAYDK 1098  
 QY 181 NKYRFEEINGBEEIAFKMRTNVSHVGOLDDIRKNPRKFVCLANDNDHNDKDAQTVKAVL 240  
 Db 1099 NKYRFEEINGBEEIAFKMRTNVSHVGOLDDIRKNPRKFVCLANDNDHNDKDAQTVKAVL 1158  
 QY 241 RDIFYESMPPIPSPQFELPREYRFLHMHELOEWRAYDKLK 281  
 Db 1159 RDIFYESMPPIPSPQFELPREYRFLHMHELOEWRAYDKLK 1199

---

RESULT 13

ID ABW0156  
 ID ABW01536 standard; protein; 1199 AA.

AC ABW0156;  
 AC ABW0156;  
 DT 15-JAN-2004 (first entry)

DE N-acetylglicosamine-1(GlcNAc)-phosphotransferase protein.  
 DE Glycoprotein; lectin; lysosomal storage disease; gastrointestinal;  
 KW N-acetylglucosamine-1-phosphotransferase; gene therapy; enzyme;  
 OS Unidentified.

XX Human protein NOV12.

XX Human; NOX; cardiomyopathy; atherosclerosis; cancer; hypertension;  
 KW diabetes; inflammation; autoimmune disorder; allergy; blood disorder;  
 KW acquired immunodeficiency syndrome; AIDS; obesity; asthma;  
 KW immunoglobulin A neuropathy; cirrhosis; arthritis; Alzheimer's disease;  
 KW Parkinson's disease; goitre; infection; stroke; muscular dystrophy;  
 KW epilepsy; wasting disorder; neurogenesis; cell differentiation;  
 KW cell proliferation; haematopoiesis; wound healing; angiogenesis;

PN US2003124653-A1.  
 XX  
 PD 03-JUL-2003.  
 XX  
 PR 21-DEC-2001; 2001US-00023890.  
 XX  
 PR 21-DEC-2001; 2001US-00023890.  
 XX  
 PA (NOVA-) NOVAZYME PHARM INC.  
 XX  
 PI Canfield WM;  
 XX  
 DR WPI; 2003-810985/76.  
 XX  
 DR N-PSDB; AAD6249.  
 XX  
 DR WPI; 2003-810985/76.  
 XX  
 PT Producing a glycoprotein with reduced complex carbohydrates by culturing the lectin resistant mammalian cell expressing the glycoprotein for treating lysosomal storage disease.

PS Claim 8; Page 10-13; 46pp; English.

CC The present invention provides a method of producing a glycoprotein having reduced complex carbohydrates by culturing the lectin resistant mammalian cell expressing the glycoprotein. The method is useful for producing a glycoprotein with reduced complex carbohydrates for treating lysosomal storage disease. The present invention is also useful in gene therapy. The present sequence is N-acetylglucosamine (GlcNAc)-phosphotransferase protein

XX Sequence 1199 AA;

Query Match 85.9%; Score 1487; DB 7; Length 1199;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-135;  
 Matches 281; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRIVNKLNKSGFTSRKVAPRHMDRTWQELDMFPEEFDKTSFHVRHS 60  
 Db 919 DTFADSLRIVNKLNKSGFTSRKVAPRHMDRTWQELDMFPEEFDKTSFHVRHS 978  
 QY 61 EDQFAFSFYIYMSAVOPLNTSQVFEVDTQSGVLSDRECFTRIHLPLS-QDLT 120  
 Db 979 EDQFAFSFYIYMSAVOPLNTSQVFEVDTQSGVLSDRECFTRIHLPLS-QDLT 1038  
 QY 121 GLEMMLINGSKMVPADTQLNNTIPPTQSYVPPNLPPTVKSLVTNCKPVTDKHKAYDK 180  
 Db 1039 GLEMMLINGSKMVPADTQLNNTIPPTQSYVPPNLPPTVKSLVTNCKPVTDKHKAYDK 1098  
 QY 181 NKYRFEEINGBEEIAFKMRTNVSHVGOLDDIRKNPRKFVCLANDNDHNDKDAQTVKAVL 240  
 Db 1099 NKYRFEEINGBEEIAFKMRTNVSHVGOLDDIRKNPRKFVCLANDNDHNDKDAQTVKAVL 1158  
 QY 241 RDIFYESMPPIPSPQFELPREYRFLHMHELOEWRAYDKLK 281  
 Db 1159 RDIFYESMPPIPSPQFELPREYRFLHMHELOEWRAYDKLK 1199

---

RESULT 14

ID ABU07381  
 ID ABU07381 standard; protein; 1459 AA.

AC ABU07381;  
 AC ABU07381;  
 DT 28-JAN-2003 (first entry)

DE Human protein NOV12.

XX Human; NOX; cardiomyopathy; atherosclerosis; cancer; hypertension;  
 KW diabetes; inflammation; autoimmune disorder; allergy; blood disorder;  
 KW acquired immunodeficiency syndrome; AIDS; obesity; asthma;  
 KW immunoglobulin A neuropathy; cirrhosis; arthritis; Alzheimer's disease;  
 KW Parkinson's disease; goitre; infection; stroke; muscular dystrophy;  
 KW epilepsy; wasting disorder; neurogenesis; cell differentiation;  
 KW cell proliferation; haematopoiesis; wound healing; angiogenesis;



CC present invention  
XX  
SQ Sequence 132 AA;

Query Match 40.7%; Score 704; DB 4; Length 132;  
Best Local Similarity 100.0%; Pred. No. 1.4e-60; Mismatches 0; Indels 0; Gaps 0;  
Matches 132; Conservative 0;

QY ||||||| MIRINSHVQGDDIRKNPKEVCLANDNIDAHKDAQTVKAWLDEYESMPISQEL 256  
Db 1 MIRINSHVQGDDIRKNPKEVCLANDNIDAHKDAQTVKAWLDEYESMPISQEL 60  
QY 257 PREVRNRFLHMHLOQEWRYDRDKLKFWMHCVLATLNTIFSFFABOLIAALKRKLPPRR 316  
Db 61 PREVRNRFLHMHLOQEWRYDRDKLKFWMHCVLATLNTIFSFFABOLIAALKRKLPPRR 120  
QY 317 IHKEASPRIRV 328  
Db 121 IHKEASPRIRV 132

Search completed: July 26, 2004, 11:09:10  
Job time : 32.2575 secs

Run on: July 26, 2004, 11:06:34 ; Search time 9.24536 Seconds  
 (without alignments) 1831.549 Million cell updates/sec

OM protein - protein search, using sw model

Title: US-10-657-280-2

Perfect score: 1731

Sequence: DTFADSLRYVNKLNSKGFGTSRKPAHMHMIRIVWOBQDNPEEFDKTHKVRS

Scoring table: BLOSUM62

Gappen 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/podata/2/iaa/5a\_Comb.pep:\*

2: /cgn2\_6/podata/2/iaa/5B\_Comb.pep:\*

3: /cgn2\_6/podata/2/iaa/6A\_Comb.pep:\*

4: /cgn2\_6/podata/2/iaa/6B\_Comb.pep:\*

5: /cgn2\_6/podata/2/iaa/PCTUS\_Comb.pep:\*

6: /cgn2\_6/podata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

### SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	1731	100.0	328	4 US-09-635-872A-2
2	1731	100.0	328	4 US-09-635-872A-2
3	1731	100.0	328	4 US-09-635-872A-2
4	1731	100.0	328	4 US-09-635-872A-2
5	1698	98.1	328	4 US-09-635-872A-2
6	1698	98.1	328	4 US-09-635-872A-2
7	1698	98.1	328	4 US-09-635-872A-2
8	1698	98.1	328	4 US-09-635-872A-2
9	402	32.2	4 US-09-633-872A-13	Sequence 2, Appli
10	402	32.2	4 US-09-636-077A-13	Sequence 2, Appli
11	402	32.2	4 US-09-636-060C-13	Sequence 2, Appli
12	402	32.2	4 US-09-986-552-13	Sequence 2, Appli
13	145	8.4	28	4 US-09-633-872A-2
14	145	8.4	28	4 US-09-633-872A-2
15	145	8.4	28	4 US-09-636-060C-27
16	145	8.4	28	4 US-09-636-060C-27
17	113.5	6.6	545	4 US-09-956-107A-10
18	101	5.8	2662	4 US-09-595-684B-31
19	100.5	5.8	259	3 US-09-467-318-137
20	100.5	5.8	259	3 US-09-467-318-137
21	100.5	5.8	259	3 US-09-468-609A-137
22	100.5	5.8	259	3 US-09-986-552-27
23	100.5	5.8	259	4 US-09-446-872A-155
24	100.5	5.8	259	4 US-09-762-227A-137
25	100.5	5.8	259	4 US-09-763-227A-155
26	100.5	5.8	259	4 US-09-763-227A-155
27	100.5	5.8	259	5 PCT-US95-01185-137

### RESULT 1

US-09-635-872A-2

Sequence 2, Application US/09635872A

Patent No. 6534300

GENERAL INFORMATION:

APPLICANT: CANFIELD, WILLIAM

TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES

FILE REFERENCE: 195613US0

CURRENT APPLICATION NUMBER: US/09/635,872A

CURRENT FILING DATE: 2000-08-10

PRIOR APPLICATION NUMBER: 60/153,831

PRIOR FILING DATE: 1999-09-14

NUMBER OF SEQ ID NOS: 52

SOFTWARE: Patentin version 3.0

SEQ ID NO: 2

LENGTH: 328

TYPE: PRT

ORGANISM: Homo sapiens

US-09-635-872A-2

Query Match Best Local Similarity 100.0%; Score 1731; DB 4; Length 328; Matches 328; Conservativeness 0; Mismatches 0; Indels 0; Gaps 0;

QY

1 DTFADSLRYVNKLNSKGFGTSRKPAHMHMIRIVWOBQDNPEEFDKTHKVRS

Db

1 DTFADSLRYVNKLNSKGFGTSRKPAHMHMIRIVWOBQDNPEEFDKTHKVRS

60

QY

61 EDMQFASFYVTLMSAVOPNLSIQVDFEVTDQSGQLSPREITLATRHELPLSQDIT

Db

61 EDMQFASFYVTLMSAVOPNLSIQVDFEVTDQSGQLSPREITLATRHELPLSQDIT

120

QY

121 GLEHMILNCSSQMLPADTQNLNPPTQESYDPNLLPPVTKSLVNCCKPVTDKHKAYKD

Db

121 GLEHMILNCSSQMLPADTQNLNPPTQESYDPNLLPPVTKSLVNCCKPVTDKHKAYKD

180

QY

181 NYCFEFIMGEELIAFMKIRINVSHVGQLDIIRKPRKRVCLNDNIDHKKDQTVAKL

Db

181 NYCFEFIMGEELIAFMKIRINVSHVGQLDIIRKPRKRVCLNDNIDHKKDQTVAKL

240

QY

241 RDYFESEMPPSQFELPREVRFLJMHQEWRAVDKUKFWHCVLALIMTIFSP

Db

241 RDYFESEMPPSQFELPREVRFLJMHQEWRAVDKUKFWHCVLALIMTIFSP

300

QY

301 ABQLIAKRKIFPQQRRHKEASPNNIRV

Db

301 ABQLIAKRKIFPQQRRHKEASPNNIRV

328

QY

AEQQLAIKRKIFPQQRRHKEASPNNIRV

328

US-09-636-077A-2  
; Sequence 2, Application US/09636077A  
; Patent No. 6337785  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM  
; TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE  
; CURRENT APPLICATION NUMBER: US/09/636,077A  
; CURRENT FILING DATE: 2000-08-10  
; PRIORITY APPLICATION NUMBER: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO: 2  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-09-636-077A-2

Query Match 100.0%; Score 1731; DB 4; Length 328;  
Best Local Similarity 100.0%; Pred. No. 9.7e-169;  
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTFADSLRYNKLN SKGFTSRKV PAIMPHEIDRITM QELDMFPBEFDKTSFHKYRHS 60  
Db 61 EDMQAFASFYIYLMASAVPLNTSQFEDVTDQSGVLSDREIRTLA TRIHEPLS QDLT 120  
Db 61 EDMQAFASFYIYLMASAVPLNTSQFEDVTDQSGVLSDREIRTLA TRIHEPLS QDLT 120

Qy 121 GLEMILNCSKMLPADTOLANNPPTQESYYPNLPPVTKSIVNCKPVTDKIHAKYDK 180  
Db 121 GLEMILNCSKMLPADTOLANNPPTQESYYPNLPPVTKSIVNCKPVTDKIHAKYDK 180

Qy 181 NKYRFEEBETAFKMRINTSHVNGOLDDIRKNPKFVCLANDNDHNHDAQTWKA VL 240  
Db 181 NKYRFEEBETAFKMRINTSHVNGOLDDIRKNPKFVCLANDNDHNHDAQTWKA VL 240

Qy 241 RDYFESMPPIPSOPELPEBRYRFLMHLOEWRAVDKLKEWPAKTFKVKFVCLANDNDHNHDAQTWKA VL 300  
Db 241 RDYFESMPPIPSOPELPEBRYRFLMHLOEWRAVDKLKEWPAKTFKVKFVCLANDNDHNHDAQTWKA VL 300

Qy 301 AEQLALKRKIPIRRRIKEASPNRIRV 328  
Db 301 AEQLALKRKIPIRRRIKEASPNRIRV 328

---

RESULT 4  
; Sequence 2, Application US/09986552  
; Patent No. 667015  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
; FILE REFERENCE: 2100890577DV  
; CURRENT APPLICATION NUMBER: US/09/986,552  
; CURRENT FILING DATE: 2001-11-09  
; PRIORITY NUMBER: 09/635,872  
; PRIORITY FILING DATE: 2000-08-10  
; PRIORITY APPLICATION NUMBER: 1999-08-15  
; PRIORITY FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO: 2  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-09-986-552-2

Query Match 100.0%; Score 1731; DB 4; Length 328;  
Best Local Similarity 100.0%; Pred. No. 9.7e-169;  
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTFADSLRYNKLN SKGFTSRKV PAIMPHEIDRITM QELDMFPBEFDKTSFHKYRHS 60  
Db 61 EDMQAFASFYIYLMASAVPLNTSQFEDVTDQSGVLSDREIRTLA TRIHEPLS QDLT 120  
Db 61 EDMQAFASFYIYLMASAVPLNTSQFEDVTDQSGVLSDREIRTLA TRIHEPLS QDLT 120

Qy 121 GLEMILNCSKMLPADTOLANNPPTQESYYPNLPPVTKSIVNCKPVTDKIHAKYDK 180  
Db 121 GLEMILNCSKMLPADTOLANNPPTQESYYPNLPPVTKSIVNCKPVTDKIHAKYDK 180

Qy 181 NKYRFEEBETAFKMRINTSHVNGOLDDIRKNPKFVCLANDNDHNHDAQTWKA VL 240  
Db 181 NKYRFEEBETAFKMRINTSHVNGOLDDIRKNPKFVCLANDNDHNHDAQTWKA VL 240

Qy 241 RDYFESMPPIPSOPELPEBRYRFLMHLOEWRAVDKLKEWPAKTFKVKFVCLANDNDHNHDAQTWKA VL 300  
Db 241 RDYFESMPPIPSOPELPEBRYRFLMHLOEWRAVDKLKEWPAKTFKVKFVCLANDNDHNHDAQTWKA VL 300

Qy 301 AEQLALKRKIPIRRRIKEASPNRIRV 328  
Db 301 AEQLALKRKIPIRRRIKEASPNRIRV 328

**RESULT 5**  
; Sequence 8, Application US/0963872A  
; Patent No. 6534300  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
; FILE REFERENCE: 195612050  
; CURRENT APPLICATION NUMBER: US/09/635, 872A  
; CURRENT FILING DATE: 2000-08-10  
; PRIORITY APPLICATION NUMBER: 60/153, 831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 8  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; US-09-635-872A-8

Query Match 98.1%; Score 1698; DB 4; Length 328;  
Best Local Similarity 97.6%; Pred. No. 2.3e-165; Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;  
Mismatches 7; Pred. No. 2.3e-165; Indels 1; Gaps 0;

Qy 1 DTFADSLRVNKLNKSGFTSRKVPAHMPAMIDRIMQELQDMFPEEFDKTSFHRHS 60  
Db 1 DTFADSLRVNKLNKSGFTSRKVPAHMPAMIDRIMQELQDMFPEEFDKTSFHRHS 60  
Qy 61 EDMQFASFYFYIYMSAVQPLNTSQVFEVDTDDQSGVLSDRGIVSDREIRTLARHPLSQQDT 120  
Db 61 EDMQFASFYFYIYMSAVQPLNTSQVFEVDTDDQSGVLSDRGIVSDREIRTLARHPLSQQDT 120  
Qy 181 NKRFEIMGEELAFKMRNTNHSWVQGOLDLDRKNPKFVCLNDNDHNKAQTVKAVL 240  
Db 181 NKRFEIMGEELAFKMRNTNHSWVQGOLDLDRKNPKFVCLNDNDHNKAQTVKAVL 240  
Qy 121 GLEMMLINCSKMLPADTOLNNIPPTQESYYDPNLPPVTKSLVNCKPVDKIKAYKD 180  
Db 121 GLEMMLINCSKMLPADTOLNNIPPTQESYYDPNLPPVTKSLVNCKPVDKIKAYKD 180  
Qy 241 RDYESMPPIPSQFELLREYRFRLMELOEWRAFRDKLKEWTHCVLATLIMFTFSFF 300  
Db 241 RDYESMPPIPSQFELLREYRFRLMELOEWRAFRDKLKEWTHCVLATLIMFTFSFF 300  
Qy 301 AEQIALAKKIKFRRRIKEASPRIV 328  
Db 301 AEQIALAKKIKFRRRIKEASPRIV 328

---

**RESULT 6**  
; Sequence 8, Application US/09636060C  
; Patent No. 6642038  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM M  
; TITLE OF INVENTION: GLCNAc PHOSPHOTRANSFERASE OF THE LYSOSOMAL TARGETING PATHWAY  
; FILE REFERENCE: 200119US0C0NT  
; CURRENT APPLICATION NUMBER: US/09/636, 060C  
; CURRENT FILING DATE: 2000-08-10  
; PRIORITY APPLICATION NUMBER: 60/153, 831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 8  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; US-09-636-060C-8

Query Match 98.1%; Score 1698; DB 4; Length 328;  
Best Local Similarity 97.6%; Pred. No. 2.3e-165; Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;  
Mismatches 7; Pred. No. 2.3e-165; Indels 1; Gaps 0;

Qy 1 DTFADSLRVNKLNKSGFTSRKVPAHMPAMIDRIMQELQDMFPEEFDKTSFHRHS 60  
Db 1 DTFADSLRVNKLNKSGFTSRKVPAHMPAMIDRIMQELQDMFPEEFDKTSFHRHS 60  
Qy 61 EDMQFASFYFYIYMSAVQPLNTSQVFEVDTDDQSGVLSDRGIVSDREIRTLARHPLSQQDT 120  
Db 61 EDMQFASFYFYIYMSAVQPLNTSQVFEVDTDDQSGVLSDRGIVSDREIRTLARHPLSQQDT 120  
Qy 121 GLEMMLINCSKMLPADTOLNNIPPTQESYYDPNLPPVTKSLVNCKPVDKIKAYKD 180  
Db 121 GLEMMLINCSKMLPADTOLNNIPPTQESYYDPNLPPVTKSLVNCKPVDKIKAYKD 180  
Qy 181 NKRFEIMGEELAFKMRNTNHSWVQGOLDLDRKNPKFVCLNDNDHNKAQTVKAVL 240  
Db 181 NKRFEIMGEELAFKMRNTNHSWVQGOLDLDRKNPKFVCLNDNDHNKAQTVKAVL 240  
Qy 241 RDYESMPPIPSQFELLREYRFRLMELOEWRAFRDKLKEWTHCVLATLIMFTFSFF 300  
Db 241 RDYESMPPIPSQFELLREYRFRLMELOEWRAFRDKLKEWTHCVLATLIMFTFSFF 300  
Qy 301 AEQIALAKKIKFRRRIKEASPRIV 328

Db 301 AEQITALRKKIFRRRIKEASPDIRV 328  
; RESULT 8  
; US-09-986-552-8  
; Sequence 8, Application US/09986552  
; Patent No. 6670165  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
; FILE REFERENCE: 215089U577DIV  
; CURRENT APPLICATION NUMBER: US/09/986, 552  
; CURRENT FILING DATE: 2001-11-09  
; PRIOR APPLICATION NUMBER: 091635, 872  
; PRIOR FILING DATE: 2000-08-10  
; PRIORITY NUMBER: 60/153, 831  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 8  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; US-09-986-552-8  
; Query Match 98.1%; Score 1698; DB 4; Length 328;  
; Best Local Similarity 97.6%; Pred. No. 2.3e-165;  
; Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;  
; Qy 1 DTFAADSLRVNKLNKNSKGFTSRKPAIMPHMDRIVMQLQEDMFPBEFDKTSFHKVRS 60  
; Db 1 DTFAADSLRVNKLNKNSKGFTSRKPAIMPHMDRIVMQLQEDMFPBEFDKTSFHKVRS 60  
; Qy 61 EDMOFAFSFYIYMSAVOPLNISQVFEDVTDQSGVLSDRREITLTATRHLPLSQDLT 120  
; Db 61 EDMOFAFSFYIYMSAVOPLNISQVFEDVTDQSGVLSDRREITLTATRHLPLSQDLT 120  
; Qy 121 GLEMMLINCSK---MUPADITOLNNLPTPOTSYDDNLPPTVSILVNCPTKSVTCKEVTDKIHKA 176  
; Db 121 GLEMMLINCSK---MUPADITOLNNLPTPOTSYDDNLPPTVSILVNCPTKSVTCKEVTDKIHKA 176  
; Qy 121 GLEMMLINCSKMLIPADITOLNNLPTPOTSYDDNLPPTVSILVNCPTKSVTCKEVTDKIHKA 180  
; Db 121 GLEMMLINCSKMLIPADITOLNNLPTPOTSYDDNLPPTVSILVNCPTKSVTCKEVTDKIHKA 180  
; Qy 181 NKRFERFEIGEEELAPKMRITVSHVQGDDIRKNPKFVCLNDNDHMHDAQTKVNL 240  
; Db 181 NKRFERFEIGEEELAPKMRITVSHVQGDDIRKNPKFVCLNDNDHMHDAQTKVNL 240  
; Qy 241 RDYYESMMPIPSQELPEYRFLRFLAMBLQWRAFDKLMKWTQHATIMPFISFF 300  
; Db 241 RDYYESMMPIPSQELPEYRFLRFLAMBLQWRAFDKLMKWTQHATIMPFISFF 300  
; Qy 301 AEQIALTRKIKPARRRKEASPNRIRV 328  
; Db 301 AEQIALTRKIKPARRRKEASPNRIRV 328  
; RESULT 9  
; US-09-635-872A-13  
; Sequence 13, Application US/09635872A  
; Patent No. 654300  
; GENERAL INFORMATION:  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
; CURRENT APPLICATION NUMBER: US/09/635, 872A  
; CURRENT FILING DATE: 2000-08-10  
; PRIORITY NUMBER: 60/153, 831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 13  
; LENGTH: 502  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
; US-09-635-872A-13  
; Query Match 23.2%; Score 402; DB 4; Length 502;  
; Best Local Similarity 34.5%; Pred. No. 1.3e-32;  
; Matches 81; Conservative 55; Mismatches 91; Indels 8; Gaps 3;  
; Qy 1 DTFAADSLRVNKLNKNSKGFTSRKPAIMPHMDRIVMQLQEDMFPBEFDKTSFHKVRS 60  
; Db 259 DIVSHLILATNMNLRAGFKARHVLTAVGFLIDKDVTEAMQRFRHQIQLTAHQFRRAP 318  
; Qy 61 EDMOFAFSFYIYMSAVOPLNISQVFEDVTDQSGVLSDRREITLTATRHLPLSQDLT 120  
; Db 319 TDQYAFAYTSFMSMETKUMSVBEIFBFDTDSATWSDRREITLTATRHLPLSQDLT 378  
; Qy 121 GLEMMLINCSK---MUPADITOLNNLPTPOTSYDDNLPPTVSILVNCPTKSVTCKEVTDKIHKA 176  
; Db 379 YFERVKVONCTRNLGMHLKUDTEBHTL--VIERYEDSNLPTITRDVRCPLAALAN 436  
; Qy 177 YDKDNKRYPEIMGR--EELAPKMRITVSHVQGDDIRKNPKFVCLNDNDH 229  
; Db 437 FAVRPKYNFHVSPTKTSNSNFMILTSNLTEVVESLDRLRNPRKFNCINDNDAN 491  
; RESULT 10  
; US-09-636-077A-13  
; Sequence 13, Application US/09636077A  
; Patent No. 6537885  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM  
; TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE  
; FILE REFERENCE: 195612USD  
; CURRENT APPLICATION NUMBER: US/09/636, 077A  
; CURRENT FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153, 831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 13  
; LENGTH: 502  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
; US-09-636-077A-13  
; Query Match 23.2%; Score 402; DB 4; Length 502;  
; Best Local Similarity 34.5%; Pred. No. 1.3e-32;  
; Matches 81; Conservative 55; Mismatches 91; Indels 8; Gaps 3;  
; Qy 1 DTFAADSLRVNKLNKNSKGFTSRKPAIMPHMDRIVMQLQEDMFPBEFDKTSFHKVRS 60  
; Db 259 DIVSHLILATNMNLRAGFKARHVLTAVGFLIDKDVTEAMQRFRHQIQLTAHQFRRAP 318  
; Qy 61 EDMOFAFSFYIYMSAVOPLNISQVFEDVTDQSGVLSDRREITLTATRHLPLSQDLT 120  
; Db 319 TDQYAFAYTSFMSMETKUMSVBEIFBFDTDSATWSDRREITLTATRHLPLSQDLT 378  
; Qy 121 GLEMMLINCSK---MUPADITOLNNLPTPOTSYDDNLPPTVSILVNCPTKSVTCKEVTDKIHKA 176  
; Db 379 YFERVKVONCTRNLGMHLKUDTEBHTL--VIERYEDSNLPTITRDVRCPLAALAN 436  
; Qy 177 YDKDNKRYPEIMGR--EELAPKMRITVSHVQGDDIRKNPKFVCLNDNDH 229  
; Db 437 FAVRPKYNFHVSPTKTSNSNFMILTSNLTEVVESLDRLRNPRKFNCINDNDAN 491  
; RESULT 11  
; US-09-636-060C-13  
; Sequence 13, Application US/09636060C  
; Patent No. 6642038  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM M  
; TITLE OF INVENTION: GLCNAC PHOSPHOTRANSFERASE OF THE LYSOSOMAL TARGETING PATHWAY  
; FILE REFERENCE: 210119US0C0NT  
; CURRENT APPLICATION NUMBER: US/09/636, 060C  
; CURRENT FILING DATE: 2000-08-10

; PRIOR APPLICATION NUMBER: 60/153,091-14  
 ; PRIOR FILING DATE: 1999-09-14  
 ; NUMBER OF SEQ ID NOS: 52  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO: 13  
 ; LENGTH: 502  
 ; TYPE: PRT  
 ; ORGANISM: Drosophila melanogaster  
 ; US-09-636-050C-13

Query Match 23.2%; Score 402; DB 4; Length 502;  
 Best Local Similarity 34.5%; Pred. No. 1.3e-32;  
 Matches 81; Conservative 55; Mismatches 91; Indels 8; Gaps 3;

QY 1 DTFADSLRVNKLNSKFGFTSRKVPAH  
 Db 259 DIVSHSLATNMNLNRAGKARHVLAVGFLIDKIVTEAMBRHQIILDRQRHQRAP 318

QY 61 EDMQFAFFYYVYIMSAVOPINTSQVDFDVTQSGVISDREITLARHPLS-QDLT 120

Db 319 TDQYQAFAYFSMSETKMSVBFDFPDTDSATWDREVTPLRIVQPLDMSMR 378

QY 121 GLEMHLINCSK---MLRADITONINTPTQBSYYDRPLPPVTSKLVSYNCKVTDKHKKA 176

Db 379 YFEEBVQNCTRNLGMHLKDVTWHTL--VYEREDSNLPTT-TRDVLRCPLAALAAAN 436

QY 177 YKOKNKPFREIMER--EELAFKOMTRNTVSHVQGQDDTRKRNPKFCNDNDHN 229

Db 437 FAVRPKKMFHVSPEERTSHSNEMMITSNTEVWSDLRURRNPKFCNDNDAN 491

RESULT 12

US-09-986-552-13  
 Sequence 13, Application US/09986552  
 ; Patent No. 6671165

; GENERAL INFORMATION:  
 ; APPLICANT: CANFIELD, WILLIAM am  
 ; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
 ; FILE REFERENCE: 215089US7DIV  
 ; CURRENT APPLICATION NUMBER: US/09/986,552  
 ; CURRENT FILING DATE: 2001-11-09  
 ; PRIORITY APPLICATION NUMBER: 09/635,872  
 ; PRIORITY FILING DATE: 2000-08-10  
 ; PRIORITY APPLICATION NUMBER: 60/153,831  
 ; PRIORITY FILING DATE: 1999-09-14  
 ; NUMBER OF SEQ ID NOS: 52  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO: 13  
 ; LENGTH: 502  
 ; TYPE: PRT  
 ; ORGANISM: Drosophila melanogaster  
 ; US-09-986-552-13

Query Match 23.2%; Score 402; DB 4; Length 502;  
 Best Local Similarity 34.5%; Pred. No. 1.3e-32;  
 Matches 81; Conservative 55; Mismatches 91; Indels 8; Gaps 3;

QY 1 DTFADSLRVNKLNSKFGFTSRKVPAH  
 Db 259 DIVSHSLATNMNLNRAGKARHVLAVGFLIDKIVTEAMBRHQIILDRQRHQRAP 318

QY 61 EDMQFAFFYYVYIMSAVOPINTSQVDFDVTQSGVISDREITLARHPLS-QDLT 120

Db 319 TDQYQAFAYFSMSETKMSVBFDFPDTDSATWDREVTPLRIVQPLDMSMR 378

QY 121 GLEMHLINCSK---MLRADITONINTPTQBSYYDRPLPPVTSKLVSYNCKVTDKHKKA 176

Db 379 YFEEBVQNCTRNLGMHLKDVTWHTL--VYEREDSNLPTT-TRDVLRCPLAALAAAN 436

RESULT 13

US-09-635-872A-27  
 Sequence 27, Application US/09635872A  
 ; Patent No. 6534300

; GENERAL INFORMATION:  
 ; APPLICANT: CANFIELD, WILLIAM  
 ; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
 ; FILE REFERENCE: 19561US0  
 ; CURRENT APPLICATION NUMBER: US/09/635,872A  
 ; CURRENT FILING DATE: 2000-08-10  
 ; PRIORITY FILING DATE: 1999-09-14  
 ; NUMBER OF SEQ ID NOS: 52  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO: 27  
 ; LENGTH: 28  
 ; TYPE: PRT  
 ; ORGANISM: Bos taurus  
 ; US-09-635-872A-27

Query Match 8.4%; Score 145; DB 4; Length 28;  
 Best Local Similarity 100.0%; Pred. No. 3.8e-08;  
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRVNKLNSKFGFTSRKVPAH 28

Db 1 DTFADSLRVNKLNSKFGFTSRKVPAH 28

RESULT 14

US-09-636-077A-27  
 Sequence 27, Application US/09636077A  
 ; Patent No. 6537785

; GENERAL INFORMATION:  
 ; APPLICANT: CANFIELD, WILLIAM  
 ; TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE  
 ; FILE REFERENCE: 195612US0  
 ; CURRENT APPLICATION NUMBER: US/09/636,077A  
 ; CURRENT FILING DATE: 2000-08-10  
 ; PRIORITY APPLICATION NUMBER: 60/153,831  
 ; PRIORITY FILING DATE: 1999-09-14  
 ; NUMBER OF SEQ ID NOS: 52  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO: 27  
 ; LENGTH: 28  
 ; TYPE: PRT  
 ; ORGANISM: Bos taurus  
 ; US-09-636-077A-27

Query Match 8.4%; Score 145; DB 4; Length 28;  
 Best Local Similarity 100.0%; Pred. No. 3.8e-08;  
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRVNKLNSKFGFTSRKVPAH 28

Db 1 DTFADSLRVNKLNSKFGFTSRKVPAH 28

RESULT 15

US-09-636-060C-27  
 Sequence 27, Application US/09636060C  
 ; Patent No. 6642038

; GENERAL INFORMATION:  
 ; APPLICANT: CANFIELD, WILLIAM M  
 ; TITLE OF INVENTION: GLCNAC PHOSPHOTRANSFERASE OF THE LYSOSOMAL TARGETING PATHWAY  
 ; FILE REFERENCE: 210119US0CONT  
 ; CURRENT APPLICATION NUMBER: US/09/636,060C  
 ; CURRENT FILING DATE: 2000-08-10  
 ; PRIORITY APPLICATION NUMBER: 60/153,831  
 ; PRIORITY FILING DATE: 1999-09-14  
 ; NUMBER OF SEQ ID NOS: 52

; SEQ ID NO 27  
; LENGTH: 28  
; TYPE: PRT  
; ORGANISM: Bos taurus  
US-09-636-060C-27

Query Match 8.4%; Score 145; DB 4; Length 28;  
Best Local Similarity 100.0%; Pred. No. 3.8e-08;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1 DTFADSLRVNKILNSKGFTSRKVPAH 28
Db	1 DTFADSLRVNKILNSKGFTSRKVPAH 28

Search completed: July 26, 2004, 11:13:19  
Job time : 10.2454 secs

Copyright (c) 1993 - 2004 Compugen Ltd.

Om protein - protein search, using sw model

Run on: July 26, 2004, 11:11:44 ; Search time 25.2146 Seconds

(without alignments) 4073.614 Million Cell updates/sec

Title: US-10-657-280-2

Perfect score: 1731

Sequence: 1 DTFADSLRYNKILNSKRGF.....RKIFPRRIKHEASPNRIRV 328

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1288442 seqs, 313154207 residues

Total number of hits satisfying chosen parameters: 1288442

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_AA:\*

1: /cgn2\_6/prodata/1/pupaa/us07\_PUBCOMB.pep:\*

2: /cgn2\_6/prodata/1/pupaa/us08\_PUBCOMB.pep:\*

3: /cgn2\_6/prodata/1/pupaa/us06\_PUBCOMB.pep:\*

4: /cgn2\_6/prodata/1/pupaa/us06\_PUBCOMB.pep:\*

5: /cgn2\_6/prodata/1/pupaa/us07\_PUBCOMB.pep:\*

6: /cgn2\_6/prodata/1/pupaa/us08\_PUBCOMB.pep:\*

7: /cgn2\_6/prodata/1/pupaa/us08\_PUBCOMB.pep:\*

8: /cgn2\_6/prodata/1/pupaa/us09\_PUBCOMB.pep:\*

9: /cgn2\_6/prodata/1/pupaa/us09\_PUBCOMB.pep:\*

10: /cgn2\_6/prodata/1/pupaa/us06\_PUBCOMB.pep:\*

11: /cgn2\_6/prodata/1/pupaa/us09\_PUBCOMB.pep:\*

12: /cgn2\_6/prodata/1/pupaa/us09\_PUBCOMB.pep:\*

13: /cgn2\_6/prodata/1/pupaa/us10\_PUBCOMB.pep:\*

14: /cgn2\_6/prodata/1/pupaa/us10\_PUBCOMB.pep:\*

15: /cgn2\_6/prodata/1/pupaa/us10\_PUBCOMB.pep:\*

16: /cgn2\_6/prodata/1/pupaa/us10\_PUBCOMB.pep:\*

17: /cgn2\_6/prodata/1/pupaa/us10\_PUBCOMB.pep:\*

18: /cgn2\_6/prodata/1/pupaa/us60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

**SUMMARIES**

Result No.	Score	Query Length	DB ID	Description
1	1731	100.0	328	9 US-09-895-072-2
2	1731	100.0	328	9 US-09-895-072-2
3	1731	100.0	328	14 US-10-023-888-5
4	1731	100.0	328	14 US-10-023-889-5
5	1731	100.0	328	14 US-10-023-890-5
6	1731	100.0	328	14 US-10-024-197-5
7	1731	100.0	328	14 US-10-024-198-4
8	1698	98.1	328	9 US-09-895-072-8
9	1698	98.1	328	9 US-09-895-072-8
10	1698	98.1	328	14 US-10-023-889-10
11	1698	98.1	328	14 US-10-023-889-10
12	1698	98.1	328	14 US-10-023-890-10
13	1698	98.1	328	14 US-10-023-890-10
14	1698	98.1	328	14 US-10-024-197-10
15	1698	98.1	328	14 US-10-024-197-10

**ALIGNMENTS**

RESULT 1

US-09-895-072-2

; Sequence 2, Application US/09895072

; Patent No. US2002005550A1

; GENERAL INFORMATION:

; APPLICANT: CANFIELD, WILLIAM M

; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES

; FILE REFERENCE: 210119US001

; CURRENT APPLICATION NUMBER: US/09/895,072

; CURRENT FILING DATE: 2001-07-02

; PRIOR APPLICATION NUMBER: 6/0153,831

; PRIOR FILING DATE: 1999-09-14

; PRIOR APPLICATION NUMBER: US 09/635,872

; PRIOR FILING DATE: 2000-08-10

; SOFTWARE: Patentin version 3.1

; SEQ ID NO 2

; LENGTH: 328

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-895-072-2

Query Match 100.0%; Score 1731; DB 9; Length 328;

Best Local Similarity 100.0%; Pred. No. 4 3e-163; Indexes 0; Gaps 0;

Matches 328; Conservative 0; Misnatches 0; Indels 0;

Sequence 2, Appli

Sequence 3, Appli

Sequence 4, Appli

Sequence 5, Appli

Sequence 6, Appli

Sequence 7, Appli

Sequence 8, Appli

Sequence 9, Appli

Sequence 10, Appli

Sequence 11, Appli

Sequence 12, Appli

Sequence 13, Appli

Sequence 14, Appli

Sequence 15, Appli

Sequence 16, Appli

Sequence 17, Appli

Sequence 18, Appli

Sequence 19, Appli

Sequence 20, Appli

Sequence 21, Appli

Sequence 22, Appli

Sequence 23, Appli

Sequence 24, Appli

Sequence 25, Appli

Sequence 26, Appli

Sequence 27, Appli

Sequence 28, Appli

Sequence 29, Appli

Sequence 30, Appli

Sequence 31, Appli

Sequence 32, Appli

Sequence 33, Appli

Sequence 34, Appli

Sequence 35, Appli

Sequence 36, Appli

Sequence 37, Appli

Sequence 38, Appli

Sequence 39, Appli

Sequence 40, Appli

Sequence 41, Appli

Sequence 42, Appli

Sequence 43, Appli

Sequence 44, Appli

Sequence 45, Appli

Sequence 46, Appli

Sequence 47, Appli

Sequence 48, Appli

Sequence 49, Appli

Sequence 50, Appli

Sequence 51, Appli

Sequence 52, Appli

Sequence 53, Appli

Sequence 54, Appli

Sequence 55, Appli

Sequence 56, Appli

Sequence 57, Appli

Sequence 58, Appli

Sequence 59, Appli

Sequence 60, Appli

Qy 1 DTFADSLRYNKILNSKRGF.....RKIFPRRIKHEASPNRIRV 60

Db 1 DTFADSLRYNKILNSKRGF.....RKIFPRRIKHEASPNRIRV 60

Qy 61 EDMQPAFSFYIYLMASQPLNTSQVFDYDVTOSGVSDREIRLARTHEPLSQLDT 120

Db 61 EDMQPAFSFYIYLMASQPLNTSQVFDYDVTOSGVSDREIRLARTHEPLSQLDT 120

Qy 121 GLEHMILNSKMLPADIQLNINPPTQESYYDPNLPPVTKSLVNCPTVKDKHCKAYDK 180

Db 121 GLEHMILNSKMLPADIQLNINPPTQESYYDPNLPPVTKSLVNCPTVKDKHCKAYDK 180

QY 181 NKRREIENGEEIAFKMIRNTSHVVGOLDDIRKNPRKFVCLNDNIDHNAKDAQTAKVL 240 ; CURRENT FILING DATE: 2001-12-21  
 Db 181 NKRREIENGEEIAFKMIRNTSHVVGOLDDIRKNPRKFVCLNDNIDHNAKDAQTAKVL 240 ; NUMBER OF SEQ ID NOS: 38  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 5  
 QY 241 RDYESMFPIPSQELPPEYRFLHMEBLQEWRAYDKLKEWTHCVLATIMFTSFF 300 ; LENGTH: 328  
 Db 241 RDYESMFPIPSQELPPEYRFLHMEBLQEWRAYDKLKEWTHCVLATIMFTSFF 300 ;  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 RESULT 2  
 US-09-986-552-2  
 ; Sequence 2, Application US/09386552  
 ; PATENT NO. US2002010981A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CAMPFIELD, William  
 ; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSMAL HYDROLASES  
 ; CURRENT APPLICATION NUMBER: US/09/986, 552  
 ; CURRENT FILING DATE: 2001-11-09  
 ; PRIOR APPLICATION NUMBER: 091635, 872  
 ; PRIOR FILING DATE: 2000-08-10  
 ; PRIOR APPLICATION NUMBER: 60/153, 831  
 ; PRIOR FILING DATE: 1999-09-14  
 ; NUMBER OF SEQ ID NOS: 52  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 2  
 ; LENGTH: 328  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-986-552-2  
 Query Match 100.0%; Score 1731; DB 9; Length 328;  
 Best Local Similarity 100.0%; Pred. No. 4.3e-163; Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DTFADSLRVNKLNKNGFTSRKVAPMMPHMIDTVQELDMFPEEFDKTSFHKVRHS 60 ;  
 Db 1 DTFADSLRVNKLNKNGFTSRKVAPMMPHMIDTVQELDMFPEEFDKTSFHKVRHS 60 ;  
 ; Publication No. US20030124652A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CAMPFIELD, William  
 ; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSMAL HYDROLASES  
 ; CURRENT APPLICATION NUMBER: US/09/986, 552  
 ; CURRENT FILING DATE: 2001-11-09  
 ; PRIOR APPLICATION NUMBER: 091635, 872  
 ; PRIOR FILING DATE: 2000-08-10  
 ; PRIOR APPLICATION NUMBER: 60/153, 831  
 ; PRIOR FILING DATE: 1999-09-14  
 ; NUMBER OF SEQ ID NOS: 52  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 2  
 ; LENGTH: 328  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-986-552-2  
 Query Match 100.0%; Score 1731; DB 14; Length 328;  
 Best Local Similarity 100.0%; Pred. No. 4.3e-163; Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DTFADSLRVNKLNKNGFTSRKVAPMMPHMIDTVQELDMFPEEFDKTSFHKVRHS 60 ;  
 Db 1 DTFADSLRVNKLNKNGFTSRKVAPMMPHMIDTVQELDMFPEEFDKTSFHKVRHS 60 ;  
 ; Publication No. US20030124652A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CAMPFIELD, William  
 ; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSMAL HYDROLASES  
 ; CURRENT APPLICATION NUMBER: US/09/986, 552  
 ; CURRENT FILING DATE: 2001-11-09  
 ; PRIOR APPLICATION NUMBER: 091635, 872  
 ; PRIOR FILING DATE: 2000-08-10  
 ; PRIOR APPLICATION NUMBER: 60/153, 831  
 ; PRIOR FILING DATE: 1999-09-14  
 ; NUMBER OF SEQ ID NOS: 52  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 2  
 ; LENGTH: 328  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-986-552-2  
 Query Match 100.0%; Score 1731; DB 14; Length 328;  
 Best Local Similarity 100.0%; Pred. No. 4.3e-163; Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DTFADSLRVNKLNKNGFTSRKVAPMMPHMIDTVQELDMFPEEFDKTSFHKVRHS 60 ;  
 Db 1 DTFADSLRVNKLNKNGFTSRKVAPMMPHMIDTVQELDMFPEEFDKTSFHKVRHS 60 ;  
 ; Publication No. US20030124652A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CAMPFIELD, William  
 ; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSMAL HYDROLASES  
 ; CURRENT APPLICATION NUMBER: US/09/986, 552  
 ; CURRENT FILING DATE: 2001-11-09  
 ; PRIOR APPLICATION NUMBER: 091635, 872  
 ; PRIOR FILING DATE: 2000-08-10  
 ; PRIOR APPLICATION NUMBER: 60/153, 831  
 ; PRIOR FILING DATE: 1999-09-14  
 ; NUMBER OF SEQ ID NOS: 52  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 2  
 ; LENGTH: 328  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-986-552-2  
 Query Match 100.0%; Score 1731; DB 14; Length 328;  
 Best Local Similarity 100.0%; Pred. No. 4.3e-163; Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DTFADSLRVNKLNKNGFTSRKVAPMMPHMIDTVQELDMFPEEFDKTSFHKVRHS 60 ;  
 Db 1 DTFADSLRVNKLNKNGFTSRKVAPMMPHMIDTVQELDMFPEEFDKTSFHKVRHS 60 ;  
 ; Publication No. US20030124652A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CAMPFIELD, William  
 ; TITLE OF INVENTION: SORBULE GLCNAC PHOSPHOTRANSFERASE  
 ; FILE REFERENCE: 203515US7  
 ; CURRENT APPLICATION NUMBER: US/10/023, 898

QY 181 NKRPEIMEEEIAFKMIRTNVSHVVGOLDDIRKAPRKFVCLNDNIDHNKAOTVKAVL 240 ; SOFTWARE: Patentin version 3.1  
Db 181 NKRPEIMEEEIAFKMIRTNVSHVVGOLDDIRKAPRKFVCLNDNIDHNKAOTVKAVL 240 ; SEQ ID NO 5  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-10-024-197-5  
RESULT 5  
; Sequence 5, Application US/10023890  
; Publication No. US20030124653A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: METHOD OF PRODUCING GLYCOPROTEINS HAVING REDUCED COMPLEX CARBOHYDRATE  
; FILE REFERENCE: 203510US77  
; CURRENT APPLICATION NUMBER: US/10/023, 890  
; CURRENT FILING DATE: 2001-12-21  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 5  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-10-023-890-5

Query Match 100.0%; Score 1731; DB 14; Length 328;  
Best Local Similarity 100.0%; Pred. No. 4.3e-163;  
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTPADSLRYNKILNSKGFTSRKPVAPMPEMDRIVMQELQDMFPEEFDKTSFHKVRHS 60  
Db 1 DTPADSLRYNKILNSKGFTSRKPVAPMPEMDRIVMQELQDMFPEEFDKTSFHKVRHS 60  
; Sequence 5, Application US/10023890  
; Publication No. US20030124653A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: MAMMALIAN CELLS  
; FILE REFERENCE: 203510US77  
; CURRENT APPLICATION NUMBER: US/10/023, 890  
; CURRENT FILING DATE: 2001-12-21  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 5  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-10-023-890-5

Query Match 100.0%; Score 1731; DB 14; Length 328;  
Best Local Similarity 100.0%; Pred. No. 4.3e-163;  
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTPADSLRYNKILNSKGFTSRKPVAPMPEMDRIVMQELQDMFPEEFDKTSFHKVRHS 60  
Db 1 DTPADSLRYNKILNSKGFTSRKPVAPMPEMDRIVMQELQDMFPEEFDKTSFHKVRHS 60  
; Sequence 5, Application US/10023890  
; Publication No. US20030124653A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: MAMMALIAN CELLS  
; FILE REFERENCE: 203510US77  
; CURRENT APPLICATION NUMBER: US/10/023, 890  
; CURRENT FILING DATE: 2001-12-21  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 5  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-10-023-890-5

Query Match 100.0%; Score 1731; DB 14; Length 328;  
Best Local Similarity 100.0%; Pred. No. 4.3e-163;  
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTPADSLRYNKILNSKGFTSRKPVAPMPEMDRIVMQELQDMFPEEFDKTSFHKVRHS 60  
Db 1 DTPADSLRYNKILNSKGFTSRKPVAPMPEMDRIVMQELQDMFPEEFDKTSFHKVRHS 60  
; Sequence 5, Application US/10023890  
; Publication No. US20030124653A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: EXPRESSION OF LYSOSMAL HYDROLASE IN CELLS EXPRESSING PRO-N-ACETYL GLUCOSAMINIDASE  
; FILE REFERENCE: 217130US77  
; CURRENT APPLICATION NUMBER: US/10/023, 894  
; CURRENT FILING DATE: 2001-12-21  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 5  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-10-023-894-5

Query Match 100.0%; Score 1731; DB 14; Length 328;  
Best Local Similarity 100.0%; Pred. No. 4.3e-163;  
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTPADSLRYNKILNSKGFTSRKPVAPMPEMDRIVMQELQDMFPEEFDKTSFHKVRHS 60  
Db 1 DTPADSLRYNKILNSKGFTSRKPVAPMPEMDRIVMQELQDMFPEEFDKTSFHKVRHS 60  
; Sequence 5, Application US/10023894  
; Publication No. US20030143669A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: ACETYLGLUCOSAMINE-1-PHOSPHODIESTER ALPHA-N-ACETYL GLUCOSAMINIDASE  
; FILE REFERENCE: 217130US77  
; CURRENT APPLICATION NUMBER: US/10/023, 894  
; CURRENT FILING DATE: 2001-12-21  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 5  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-10-023-894-5

Query Match 100.0%; Score 1731; DB 14; Length 328;  
Best Local Similarity 100.0%; Pred. No. 4.3e-163;  
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTPADSLRYNKILNSKGFTSRKPVAPMPEMDRIVMQELQDMFPEEFDKTSFHKVRHS 60  
Db 1 DTPADSLRYNKILNSKGFTSRKPVAPMPEMDRIVMQELQDMFPEEFDKTSFHKVRHS 60  
; Sequence 5, Application US/10023894  
; Publication No. US20030143669A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: HIGHLY PHOSPHORYLATED ACID BETA-GLUCOCERBROSIDASE AND METHODS OF TREATMENT  
; FILE REFERENCE: 209794US0  
; CURRENT APPLICATION NUMBER: US/10/024, 197  
; CURRENT FILING DATE: 2001-12-21  
; NUMBER OF SEQ ID NOS: 27

RESULT 8

```

; PRIORITY INFORMATION:
;   PRIORITY NUMBER: 60/153,831
;   PRIORITY DATE: 1999-09-14
;   PRIOR APPLICATION NUMBER: US 09/635,872
;   PRIOR FILING DATE: 2000-08-10
;   SOFTWARE: PatentIn version 3.1
;   SEQ ID NO: 8
;   LENGTH: 328
;   TYPE: PRT
;   ORGANISM: Mus musculus
;   US-09-895-072-8

Query Match 98.1%; Score 1698; DB 9; Length 328;
Best Local Similarity 97.6%; Pred. No. 8e-160; 1; Indels 0; Gaps 0;
Matches 320; Conservative 7; Mismatches 1; Index 0; Gaps 0;

Query 1 DTFAASLRVUNKLNSKGFTSKRVPAMPHMIDRIVQELDMFPEEFDKTSFHKYRS 60
Db 1 DTFAASLRVUNKLNSKGFTSKRVPAMPHMIDRIVQELDMFPEEFDKTSFHKYRS 60
Qy 61 EDMQFAFSYFYIYMSAVOPLNISQFDVTDQSGVLSDREIRTLAIREHPLS1QDLT 120
Db 61 EDMQFAFSYFYIYMSAVOPLNISQFDVTDQSGVLSDREIRTLAIREHPLS1QDLT 120
Qy 121 GLEMNLINCSKMLPANTIQNNTPPQTDAYDPNLPPVTKSLVTNCPKVTDKHYKDK 180
Db 121 GLEMNLINCSKMLPANTIQNNTPPQTDAYDPNLPPVTKSLVTNCPKVTDKHYKDK 180
Qy 181 NKYREFEINGEEETAFKMRTNVSHVGQDDTRKNPKFVCLANDNDHNHDQATPVKAVL 240
Db 181 NKYREFEINGEEETAFKMRTNVSHVGQDDTRKNPKFVCLANDNDHNHDQATPVKAVL 240
Qy 241 RDYESMFP1PSOELPESEYRFLHMELOEWRAYDKLKEWTHCVLATIMFTFSFF 300
Db 241 RDYESMFP1PSOELPESEYRFLHMELOEWRAYDKLKEWTHCVLATIMFTFSFF 300
Qy 301 AEQIALALKKIFRRRIKEASPNRIRV 328
Db 301 AEQIALALKKIFRRRIKEASPNRIRV 328

RESULT 9
; Sequence 2, Application US/10306686
; Publication No. US20030148460A1
; GENERAL INFORMATION:
;   APPLICANT: CANFIELD, WILLIAM
;   TITLE OF INVENTION: PHOSPHODISTER ALPHA-GLCNAcase OF THE LYSOSMAL TARGETING PATHWAY
;   FILE REFERENCE: 210397US77DIV
;   CURRENT APPLICATION NUMBER: US/10/306, 686
;   CURRENT FILING DATE: 2002-11-29
;   PRIOR APPLICATION NUMBER: 09/636, 596
;   PRIOR FILING DATE: 2000-08-10
;   PRIORITY NUMBER: 60/153, 831
;   PRIORITY FILING DATE: 1999-08-14
;   NUMBER OF SEQ ID NOS: 52
;   SOFTWARE: PatentIn version 3.1
;   SEQ ID NO: 2
;   LENGTH: 328
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   US-10-306-686-2

Query Match 100.0%; Score 1731; DB 14; Length 328;
Best Local Similarity 100.0%; Pred. No. 4.3e-163; 0; Mismatches 0; Index 0; Gaps 0;
Matches 328; Conservative 0; Mismatches 0; Index 0; Gaps 0;

Query 1 DTFAASLRVUNKLNSKGFTSKRVPAMPHMIDRIVQELDMFPEEFDKTSFHKYRS 60
Db 1 DTFAASLRVUNKLNSKGFTSKRVPAMPHMIDRIVQELDMFPEEFDKTSFHKYRS 60
Qy 61 EDMQFAFSYFYIYMSAVOPLNISQFDVTDQSGVLSDREIRTLAIREHPLS1QDLT 120
Db 61 EDMQFAFSYFYIYMSAVOPLNISQFDVTDQSGVLSDREIRTLAIREHPLS1QDLT 120
Qy 121 GLEMNLINCSKMLPANTIQNNTPPQTDAYDPNLPPVTKSLVTNCPKVTDKHYKDK 180
Db 121 GLEMNLINCSKMLPANTIQNNTPPQTDAYDPNLPPVTKSLVTNCPKVTDKHYKDK 180
Qy 181 NKYREFEINGEEETAFKMRTNVSHVGQDDTRKNPKFVCLANDNDHNHDQATPVKAVL 240
Db 181 NKYREFEINGEEETAFKMRTNVSHVGQDDTRKNPKFVCLANDNDHNHDQATPVKAVL 240
Qy 241 RDYESMFP1PSOELPESEYRFLHMELOEWRAYDKLKEWTHCVLATIMFTFSFF 300
Db 241 RDYESMFP1PSOELPESEYRFLHMELOEWRAYDKLKEWTHCVLATIMFTFSFF 300
Qy 301 AEQIALALKKIFRRRIKEASPNRIRV 328
Db 301 AEQIALALKKIFRRRIKEASPNRIRV 328

RESULT 10
; Sequence 8, Application US/09986552
; Patent No. US2002150981A1
; GENERAL INFORMATION:
;   APPLICANT: CANFIELD, William
;   TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSMAL HYDROLASES
;   FILE REFERENCE: 21089U577DIV
;   CURRENT APPLICATION NUMBER: US/09/986, 552
;   PRIORITY NUMBER: 09/635, 872
;   PRIOR FILING DATE: 2000-08-10
;   PRIOR APPLICATION NUMBER: 60/153, 831
;   PRIOR FILING DATE: 1999-09-14
;   NUMBER OF SEQ ID NOS: 52
;   SOFTWARE: PatentIn version 3.1
;   SEQ ID NO: 8
;   LENGTH: 328
;   TYPE: PRT
;   ORGANISM: Mus musculus
;   US-09-986-552-8

Query Match 98.1%; Score 1698; DB 9; Length 328;
Best Local Similarity 97.6%; Pred. No. 8e-160; 1; Indels 0; Gaps 0;
Matches 320; Conservative 7; Mismatches 1; Index 0; Gaps 0;

Query 1 DTFAASLRVUNKLNSKGFTSKRVPAMPHMIDRIVQELDMFPEEFDKTSFHKYRS 60
Db 1 DTFAASLRVUNKLNSKGFTSKRVPAMPHMIDRIVQELDMFPEEFDKTSFHKYRS 60
Qy 61 EDMQFAFSYFYIYMSAVOPLNISQFDVTDQSGVLSDREIRTLAIREHPLS1QDLT 120
Db 61 EDMQFAFSYFYIYMSAVOPLNISQFDVTDQSGVLSDREIRTLAIREHPLS1QDLT 120

```

Db 61 EDMQAFASFYVYIMSAVOPLNISQVHEVDTDQSGVLSDRBIRTLATRIDLPLSLQDLT 120 ; TITLE OF INVENTION: DEFICIENT CELLS  
 Qy 121 GLEHMILNCSKMLPADITOLNNIPTQESYYDPNLPPVTSLVNTCKPVTDKIHAYKK 180 ; FILE REFERENCE: 203512U577  
 Db 121 GLEHMILNCSKMLPADITOLNNIPTQESYYDPNLPPVTSLVNTCKPVTDKIHAYKK 180 ; CURRENT APPLICATION NUMBER: US/10/023, 889  
 ; CURRENT FILING DATE: 2001-12-21  
 ; NUMBER OF SEQ ID NOS: 21  
 Qy 181 NKYRFELMGEESTAFKMRTRNVSHVQGOLDIRKPKFCLNDNDHMKDAQTVKAVL 240 ; SOFTWARE: Patentin version 3.1  
 Db 181 NKYRFELMGEESTAFKMRTRNVSHVQGOLDIRKPKFCLNDNDHMKDAQTVKAVL 240 ; SEQ ID NO 10  
 ; LENGTH: 328  
 ; TYPE: PRT  
 ; ORGANISM: Mus musculus  
 Qy 241 RDIFYESMPFIPSQFELPREYRNFLAMELOEWRAYRDKLFKWTCHVCLATLIFTIFSPF 300 ; US-10-023-889-10  
 Db 241 RDIFYESMPFIPSQFELPREYRNFLAMELOEWRAYRDKLFKWTCHVCLATLIFTIFSPF 300 ; Query Match 98.1%; Score 1698; DB 14; Length 328;  
 Qy 301 AEQIALRKKIFPRRRRIKEASPDRIV 328 ; Best Local Similarity 97.6%; Pred. No. 8e-160; 1; Mismatches 0; Indels 0; Gaps 0;  
 Db 301 AEQIALRKKIFPRRRRIKEASPDRIV 328 ; Matches 320; Conservative 7; Misnatches 0;  
 RESULT 11 ; Publication No. US20030119088A1.  
 ; Sequence 10, Application US/10023888  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CANFIELD, William  
 ; TITLE OF INVENTION: SOLUBLE GLCNAC PHOSPHOTRANSFERASE  
 ; FILE REFERENCE: 203515U577  
 ; CURRENT APPLICATION NUMBER: US/10/023, 888  
 ; CURRENT FILING DATE: 2001-12-21  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SOFTWARE: Patentin version 3.1  
 ; SEQ ID NO 10  
 ; LENGTH: 328  
 ; TYPE: PRT  
 ; ORGANISM: Mus musculus  
 US-10-023-888-10  
 Query Match 98.1%; Score 1698; DB 14; Length 328;  
 Best Local Similarity 97.6%; Pred. No. 8e-160; 1; Mismatches 0; Indels 0; Gaps 0;  
 Matches 320; Conservative 7; Misnatches 0;  
 Qy 1 DTFADSLRYVNKLNSKGFTSRKVPAHMPMIDRIVQELDMFPEEFDKTSFHKVRS 60 ; Query Match 98.1%; Score 1698; DB 14; Length 328;  
 Db 1 DTFADSLRYVNKLNSKGFTSRKVPAHMPMIDRIVQELDMFPEEFDKTSFHKVRS 60 ; Best Local Similarity 97.6%; Pred. No. 8e-160; 1; Mismatches 0; Indels 0; Gaps 0;  
 Qy 61 EDMQAFASFYVYIMSAVOPLNISQVHEVDTDQSGVLSDREIRTLATRIDLPLSLQDLT 120 ; Matches 320; Conservative 7; Misnatches 0;  
 Db 61 EDMQAFASFYVYIMSAVOPLNISQVHEVDTDQSGVLSDREIRTLATRIDLPLSLQDLT 120 ; Query Match 98.1%; Score 1698; DB 14; Length 328;  
 Qy 121 GLEHMILNCSKMLPADITOLNNIPTQESYYDPNLPPVTSLVNTCKPVTDKIHAYKK 180 ; Best Local Similarity 97.6%; Pred. No. 8e-160; 1; Mismatches 0; Indels 0; Gaps 0;  
 Db 121 GLEHMILNCSKMLPADITOLNNIPTQESYYDPNLPPVTSLVNTCKPVTDKIHAYKK 180 ; Matches 320; Conservative 7; Misnatches 0;  
 Qy 181 NKYRFELMGEESTAFKMRTRNVSHVQGOLDIRKPKFCLNDNDHMKDAQTVKAVL 240 ; Query Match 98.1%; Score 1698; DB 14; Length 328;  
 Db 181 NKYRFELMGEESTAFKMRTRNVSHVQGOLDIRKPKFCLNDNDHMKDAQTVKAVL 240 ; Best Local Similarity 97.6%; Pred. No. 8e-160; 1; Mismatches 0; Indels 0; Gaps 0;  
 Qy 241 RDIFYESMPFIPSQFELPREYRNFLAMELOEWRAYRDKLFKWTCHVCLATLIFTIFSPF 300 ; Matches 320; Conservative 7; Misnatches 0;  
 Db 241 RDIFYESMPFIPSQFELPREYRNFLAMELOEWRAYRDKLFKWTCHVCLATLIFTIFSPF 300 ; Query Match 98.1%; Score 1698; DB 14; Length 328;  
 Qy 301 AEQIALRKKIFPRRRRIKEASPDRIV 328 ; Best Local Similarity 97.6%; Pred. No. 8e-160; 1; Mismatches 0; Indels 0; Gaps 0;  
 Db 301 AEQIALRKKIFPRRRRIKEASPDRIV 328 ; Matches 320; Conservative 7; Misnatches 0;  
 RESULT 12 ; Publication No. US20030124653A1.  
 ; Sequence 10, Application US/10023890  
 ; General Information:  
 ; APPLICANT: CANFIELD, William  
 ; TITLE OF INVENTION: METHOD OF PRODUCING GLYCOPROTEINS HAVING REDUCED COMPLEX CARBOHY  
 ; FILE REFERENCE: 203510U577  
 ; CURRENT APPLICATION NUMBER: US/10/023, 890  
 ; CURRENT FILING DATE: 2001-12-21  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: Patentin version 3.1  
 ; SEQ ID NO 10  
 ; LENGTH: 328  
 ; TYPE: PRT  
 ; ORGANISM: Mus musculus  
 US-10-023-890-10  
 Query Match 98.1%; Score 1698; DB 14; Length 328;  
 Best Local Similarity 97.6%; Pred. No. 8e-160; 1; Mismatches 0; Indels 0; Gaps 0;  
 Matches 320; Conservative 7; Misnatches 0;  
 Qy 1 DTFADSLRYVNKLNSKGFTSRKVPAHMPMIDRIVQELDMFPEEFDKTSFHKVRS 60 ; Query Match 98.1%; Score 1698; DB 14; Length 328;  
 Db 1 DTFADSLRYVNKLNSKGFTSRKVPAHMPMIDRIVQELDMFPEEFDKTSFHKVRS 60 ; Best Local Similarity 97.6%; Pred. No. 8e-160; 1; Mismatches 0; Indels 0; Gaps 0;  
 Qy 61 EDMQAFASFYVYIMSAVOPLNISQVHEVDTDQSGVLSDREIRTLATRIDLPLSLQDLT 120 ; Matches 320; Conservative 7; Misnatches 0;  
 Db 61 EDMQAFASFYVYIMSAVOPLNISQVHEVDTDQSGVLSDREIRTLATRIDLPLSLQDLT 120 ; Query Match 98.1%; Score 1698; DB 14; Length 328;  
 Qy 121 GLEHMILNCSKMLPADITOLNNIPTQESYYDPNLPPVTSLVNTCKPVTDKIHAYKK 180 ; Best Local Similarity 97.6%; Pred. No. 8e-160; 1; Mismatches 0; Indels 0; Gaps 0;  
 Db 121 GLEHMILNCSKMLPADITOLNNIPTQESYYDPNLPPVTSLVNTCKPVTDKIHAYKK 180 ; Matches 320; Conservative 7; Misnatches 0;  
 Qy 181 NKYRFELMGEESTAFKMRTRNVSHVQGOLDIRKPKFCLNDNDHMKDAQTVKAVL 240 ; Query Match 98.1%; Score 1698; DB 14; Length 328;  
 Db 181 NKYRFELMGEESTAFKMRTRNVSHVQGOLDIRKPKFCLNDNDHMKDAQTVKAVL 240 ; Best Local Similarity 97.6%; Pred. No. 8e-160; 1; Mismatches 0; Indels 0; Gaps 0;  
 Qy 241 RDIFYESMPFIPSQFELPREYRNFLAMELOEWRAYRDKLFKWTCHVCLATLIFTIFSPF 300 ; Matches 320; Conservative 7; Misnatches 0;  
 Db 241 RDIFYESMPFIPSQFELPREYRNFLAMELOEWRAYRDKLFKWTCHVCLATLIFTIFSPF 300 ; Query Match 98.1%; Score 1698; DB 14; Length 328;  
 Qy 301 AEQIALRKKIFPRRRRIKEASPDRIV 328 ; Best Local Similarity 97.6%; Pred. No. 8e-160; 1; Mismatches 0; Indels 0; Gaps 0;  
 Db 301 AEQIALRKKIFPRRRRIKEASPDRIV 328 ; Matches 320; Conservative 7; Misnatches 0;

; FILE REFERENCE: 217139IS7  
; CURRENT APPLICATION NUMBER: US/10/023, 894  
; CURRENT FILING DATE: 2001-12-21  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 10  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; US-10-023-894-10

Query Match  
Best Local Similarity 98.1%; Score 1698; DB 14; Length 328;  
Matches 320; Conservatv 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 DTFADSLRYVNKLNSKGFTSRKPAHMIDRIVNQELQDMPPERPKTSFKHRS 60  
Db 1 DTFADSLRYVNKLNSKGFTSRKPAHMIDRIVNQELQDMPPERPKTSFKHRS 60

QY 61 EDMQFAFSYFYIIMSAVQPLNISQFEDVTDQSGVLSDREITLATHPLSLQDLT 120  
Db 61 EDMQFAFSYFYIIMSAVQPLNISQFEDVTDQSGVLSDREITLATHPLSLQDLT 120

QY 121 GEHMILINCSKMLPADTOLNNIPPOBYYDNPNUPTVKSLVTKPVDIKHKAYDK 180  
Db 121 GEHMILINCSKMLPADTOLNNIPPOBYYDNPNUPTVKSLVTKPVDIKHKAYDK 180

QY 181 NKYRFELMEGEEIAFKMIRTNVSHVGOLDIIRKPRKFVCLNDNDHMKDQTYKAVL 240  
Db 181 NKYRFELMEGEEIAFKMIRTNVSHVGOLDIIRKPRKFVCLNDNDHMKDQTYKAVL 240

QY 241 RDPYESMPPIPSQFELPREYRNFLAMELQEWRAYDRDKLKFWTHCVLATLIFTIFSPF 300  
Db 241 RDPYESMPPIPSQFELPREYRNFLAMELQEWRAYDRDKLKFWTHCVLATLIFTIFSPF 300

QY 301 AEQIALRKIKPFRRIKEASPRIRV 328  
Db 301 AEQIALRKIKPFRRIKEASPRIRV 328

RESULT 14  
US-10-024-197-10  
; Sequence 10, Application US/10024197  
; GENERAL INFORMATION:  
; Publication No. US20030113924A1  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: HIGHLY PHOSPHORYLATED ACID BETA-GLUCOCEREBROSIDASE AND METHODS OF TREATMENT FOR GAUCHER'S DISEASE  
; FILE REFERENCE: 209194US0  
; CURRENT APPLICATION NUMBER: US/10/024, 197  
; CURRENT FILING DATE: 2001-12-21  
; NUMBER OF SEQ ID NOS: 27  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 10  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; US-10-024-197-10

Query Match  
Best Local Similarity 97.6%; Score 1698; DB 14; Length 328;  
Matches 320; Conservatv 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 DTFADSLRYVNKLNSKGFTSRKPAHMIDRIVNQELQDMPPERPKTSFKHRS 60  
Db 1 DTFADSLRYVNKLNSKGFTSRKPAHMIDRIVNQELQDMPPERPKTSFKHRS 60

QY 61 EDMQFAFSYFYIIMSAVQPLNISQFEDVTDQSGVLSDREITLATHPLSLQDLT 120  
Db 61 EDMQFAFSYFYIIMSAVQPLNISQFEDVTDQSGVLSDREITLATHPLSLQDLT 120

QY 121 GEHMILINCSKMLPADTOLNNIPPOBYYDNPNUPTVKSLVTKPVDIKHKAYDK 180  
Db 121 GEHMILINCSKMLPADTOLNNIPPOBYYDNPNUPTVKSLVTKPVDIKHKAYDK 180

QY 181 NKYRFELMEGEEIAFKMIRTNVSHVGOLDIIRKPRKFVCLNDNDHMKDQTYKAVL 240  
Db 181 NKYRFELMEGEEIAFKMIRTNVSHVGOLDIIRKPRKFVCLNDNDHMKDQTYKAVL 240

QY 241 RDPYESMPPIPSQFELPREYRNFLAMELQEWRAYDRDKLKFWTHCVLATLIFTIFSPF 300  
Db 241 RDPYESMPPIPSQFELPREYRNFLAMELQEWRAYDRDKLKFWTHCVLATLIFTIFSPF 300

QY 301 AEQIALRKIKPFRRIKEASPRIRV 328  
Db 301 AEQIALRKIKPFRRIKEASPRIRV 328

RESULT 15  
US-10-023-894-10  
; Sequence 10, Application US/10023894  
; GENERAL INFORMATION:  
; Publication No. US20030143669A1  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: EXPRESSION OF LYSOSMAL HYDROLASE IN CELLS EXPRESSING PRO-N-ACETYL GLUCOSAMINIDASE  
; TITLE OF INVENTION: ACETYLGLUCOSAMINE-1-PHOSPHODIESTER ALPHA-N-ACETYL GLUCOSAMINIDASE

Copyright GenCore version 5.1.6  
(c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2004, 11:01:58 ; Search time 5.46316 Seconds  
3126.212 Million cell updates/sec  
(w/o alignments)

Title: US-10-657-280-2

Perfect score: 1731

Sequence: 1 DTIADSLRIVNKILNSKRGF.....RKIFPRRIKHEASPNRIRV 328

## Scoring table: BLASTm62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 10%  
Listing first 45 summaries

Database : SwissProt\_42;\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	12.4	7.2	1312	RAS0 YEAST
2	10.5	6.2	971	OPIAL_OENOMA
3	10.6	6.1	574	EZRAS_STRPY
4	10.4	6.0	574	EZRAS_STRP3
5	10.1	5.8	574	EZRAS_STRP8
6	10.1	5.8	2653	CENE_BERHAN
7	9.9	5.7	587	TATR_NPYEM
8	9.9	5.7	385	GBA2_SOYEN
9	9.7	5.6	1961	NMYB9_RAT
10	9.6	5.5	515	VAC1_YEAST
11	9.5	5.5	2867	RBP2_PLAVB
12	9.4	5.4	1	AKM0_YEAST
13	9.4	5.4	817	IGB_YEAST
14	9.4	5.4	1978	MYB_HICK
15	9.3	5.4	353	SYN_BORLU
16	93.5	5.4	1	MNYB9_HUMAN
17	91.5	5.3	1960	YN04_YEAST
18	91.5	5.3	587	1_VATA_TRICO
19	91.5	5.3	1006	1_DPOL_VACC
20	91.5	5.3	1006	1_DPOL_VACC
21	91.5	5.3	582	1_TAPR_NPYAC
22	91	5.3	1157	1_DD77_HUMAN
23	90	5.2	734	1_RELB_BASU
24	90	5.2	1	CAPP_STRPY
25	89.5	5.2	460	1_YSA1_CAREL
26	89.5	5.2	1005	1_DPOL_VARV
27	89.5	5.2	1701	1_MSPL_PLAFFF
28	89.5	5.2	1701	1_MSPL_PLAFFF
29	89	5.1	420	1_NC02_RAT
30	89	5.1	879	1_RAS0_SUITO
31	89	5.1	1013	1_PRML_DRONE
32	89	5.1	2104	1_MYC3_SCIP0
33	88.5	5.1	463	1_POF2_SCIP0

RESULT	1	RA50 YEAST	STANDARD;	PRT;	1312 AA.
ID	RA50 YEAST				
AC	P12753;				
DT	01-OCT-1989 (Rel. 12, Created)				
DT	01-OCT-1998 (Rel. 12, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	DNA repair protein RAD50 (153 kDa protein).				
GN	RAD50 OR YNL250W OR N9872;				
OS	Saccharomyces cerevisiae (Baker's yeast).				
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
OC	Saccharomycetales; Saccharomycetaceae; Saccharomyces;				
OX	NCBI_TaxID=4932;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	STRAIN=N2821; PubMed=24559437;				
RX	RENAME=228C / FY1679;				
RX	Medline=9377992; PubMed=9934673;				
RA	Sen-Gupta M., Gueldenauer U., Beinhauer J.D., Fiedler T.A.,				
RA	Heermann J.H.;				
RT	"Sequence analysis of the 33 kb long region between ORC5 and SUL1 from the left arm of chromosome XIV from <i>Saccharomyces cerevisiae</i> "; Yeast 13:849-860(1997).				
RT	- FUNCTION: Involved in DNA double-strand break repair (DSBR). The rad50/mre11 complex possesses single-strand endonuclease activity and ATP-dependent double-strand-specific exonuclease activity. Rad50 provides an ATP-dependent control of mre11 by unwinding and/or repositioning DNA ends into the mre11 active site.				
RT	--!- SUBUNIT: Forms a complex with MRE11.				
CC	This SWISS PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce">http://www.isb-sib.ch/announce</a> or send an email to license@isb-sib.ch).				
CC	DR: EMBL; X14814; CA32919; 1; -;				
DR	EMBL; X56722; CA65494; 1; -;				
DR	EMBL; Z71526; CA96157; 1; -;				
DR	PIR; S05808; BWBYD1;				
DR	Germontine; 143256; -;				
DR	SGD; S00519; RAD50;				
DR	DR: GO:0005634; C:nucleus; IDA.				
GO	GO:0003660; F:double-stranded DNA binding; IDA.				
GO	GO:0006303; P:double-strand break repair via nonhomologous				
GO	IMP.				

08K873 streptococc  
P03965 saccharomyce  
P12270 homo sapien  
Q92x3 c bifunctional  
Qcnw5 staphylococ  
P1145 trypanosoma  
P40418 kluveromyce  
Qusi6 schizosacch  
Q1923 sorghum bic  
Q4314 homo sapien  
Q99b04 buchnera ap



OC Streptococcub.  
 OX NCBI\_TaxID:1314;  
 RP SEQUENCE FROM N.A.  
 RN [1] STRAIN=SF370 / ATCC 700294 / Serotype M1;  
 RX MEDLINE=21192684; PubMed=11296296;  
 RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,  
 RA Oian Y., Jia H.G., Najar P.Z., Ren Q., Zhu H., Song L., White J.,  
 RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.,  
 RT "Complete genome sequence of an M1 strain of *Streptococcus pyogenes*."  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:658-663(2001).  
 CC -!- FUNCTION: Negative regulator of ftsZ ring formation; modulates the  
 frequency and position of ftsZ ring formation. Inhibits ftsZ ring  
 formation at polar sites. Interacts either with ftsZ or with one  
 of its binding partners to promote depolymerization (By  
 similarity).  
 -!- SUBCELLULAR LOCATION: Membrane-associated. Colocalized with ftsZ to  
 the nascent septal site (By similarity).  
 CC -!- SIMILARITY: Belongs to the eza family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AB005524; AAK3678.1; -.  
 DR HAMAP; MF 00728; -; 1.  
 KW Cell division; Septation; Transmembrane; Coiled coil;  
 KW Complete proteome.  
 FT DOMAIN 1 7 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 8 26 POTENTIAL.  
 FT DOMAIN 27 574 CTOPLASMIC (POTENTIAL).  
 FT DOMAIN 102 148 Coiled coil (Potential).  
 FT DOMAIN 276 366 Coiled coil (Potential).  
 SQ SEQUENCE 574 AA; 66120 MW; 91FP877351868827 CRC64;

Query Match 6.1%; Score 106; DB 1; Length 574;  
 Best Local Similarity 21.8%; Pred. No. 1..3; Matches 61; Conservative 48; Mismatches 113; Indels 58; Gaps 12;

OY 6 SLRVNKILNSKGFTSRKVPAHMPHMIDRTWQELQDMFPE--EFDKTSFHKVRHSD 62  
 ::|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:  
 126 SIREALNLIKEQ---BEKNSARVTHAID--LYEKLUQASISENEDNGSTMPIDKQMN 179  
 DR EMBL; AB005545; BAC6472.1; -.  
 DR HAMAP; MF 00728; -; 1.  
 KW Cell division; Septation; Transmembrane; Coiled coil;  
 KW Complete proteome.

OY 63 MOPAFSYFYIIMSAVQPLNISQVF--EVTDOSGMSDREIRTLATRHEPLSQLD- 119  
 ::|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:  
 180 IIEFSPQVALNSGDPVEASEVLDRAEBHTALGOITPAVAKLEDDPPQDIDR 239  
 IIEFSPQVALNSGDPVEASEVLDRAEBHTALGOITPAVAKLEDDPPQDIDR 239

OY 120 TGLEMMLINCSKMLPADITQLNNIPPTQESTYDP----NLPPVTSLVNCKP-VTK 172  
 ::|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:  
 240 TGYRRL-----EENTHPEKNEARFOETRESIRRANSSELVTD 279

OY 173 IHKAYKOKKNKYR-----FEIPEGEEETAFKMIRTNVSHVGQDDIKRDRPKFWCLNDN; 226  
 ::|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:  
 280 LDRAEENTHPEKNEARFOETRESIRRANSSELVTD 336

Db 227 DH-----NHKDQTVKAVRDYF--SMPPIQSQEL 256

Db 337 ARLSRKVILSERESLTWKAPEKDKIETDSTLAVAEQFGI 376

Query Match 6.0%; Score 104; DB 1; Length 574;  
 Best Local Similarity 22.5%; Pred. No. 1..9; Matches 64; Conservative 48; Mismatches 106; Indels 66; Gaps 14;

OY 6 SLRVNKILNSKGFTSRKVPAHMPHMIDRTWQELQDMFPE--EFDKTSFHKVRHSD 62  
 ::|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:  
 126 SIREALNLIKEQ---BEKNSARVTHAID--LYEKLUQASISENEDNGSTMPIDKQMN 179  
 DR EMBL; AB005545; BAC6472.1; -.  
 DR HAMAP; MF 00728; -; 1.  
 KW Cell division; Septation; Transmembrane; Coiled coil;  
 KW Complete proteome.

OY 63 MOPAFSYFYIIMSAVQPLNISQVF--EVTDOSGMSDREIRTLATRHEPLSQLD- 119  
 ::|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:  
 180 IIEFSPQVALNSGDPVEASEVLDRAEBHTALGOITPAVAKLEDDPPQDIDR 239  
 IIEFSPQVALNSGDPVEASEVLDRAEBHTALGOITPAVAKLEDDPPQDIDR 239

OY 120 TGLEMMLINCSKMLPADITQLNNIPPTQESTYDP----NLPPVTSLVNCKP-VTK 172  
 ::|||:|||:|||:|||:|||:|||:|||:|||:|||:  
 240 TGYRRL-----EENTHPEKNEARFOETRESIRRANSSELVTD 279

Septation ring formation regulator eza.

RESULT 4  
 EZRA STRP3 STANDARD; PRT; 574 AA.  
 ID EZRA STRP3 STANDARD; PRT; 574 AA.  
 AC 08K639;  
 DT 15-MAR-2004 (Rel. 43, Created)  
 DT 15-MAR-2004 (Rel. 43, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE

QY 173 IHKAYDKDKYR----PEINGEEETAFKMIRTN-----VSHTV---GQL-DDTRKN 215  
 QY :|: ;|: ;|: ;|: ;|: ;|: ;|: ;|: ;|: ;|: ;|: ;|: ;|: ;|: ;|: ;|: ;|: ;|:  
 QY 280 LDRAREENTHIQRIDSLYEVFERELAYKVAKRNQMLPRLAHYKRNNNEQLKDIARL 339  
 QY 216 PRKFVCLNDNIDINHKDQTVKAVALDYE--SMFPIPSQEL 256  
 QY ||: ;|: ;|: ;|: ;|: ;|: ;|: ;|: ;|: ;|: ;|: ;|: ;|: ;|:  
 QY 340 SRKVYL-----SETESLTIVKAPEKOKIETEDSTLVAEQQGL 376  
 QY RESULT 5  
 ERRA\_STRPB STANDARD; PRT; 574 AA.  
 ID ERRA\_STRPB AC  
 ID QSPIM3; AC  
 DT 15-MAR-2004 (Rel. 43, Last sequence update)  
 DE Septation ring formation regulator erra.  
 GN EZRA OR SPYMB 0796  
 OS Streptococcus pyogenes (serotype M18).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 Spp. Seroprococcus.  
 OK NEBI\_TaxID=186103;  
 RN [1]  
 RP STRAIN=MGSAS23 / Serotype M18;  
 RX MEDLINE=21927593; Published=1991/10/08;  
 RA Shoot J.C., Barbisan R., Van Gompel J.J., Shoot L.M., Chausee M.S.,  
 RA Sylvia G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,  
 RA Parkins L.D., Beers S.B., Campbell D.S., Smith T.M., Zhang Q.,  
 RA Kapur V., Daly J.A., Veasy L.G., Muster J.M.;  
 RT "Genome sequence and comparative microarray analysis of serotype M18  
 RT group A Streptococcus strains associated with acute rheumatic fever  
 RT outbreaks.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4658-4673 (2002).  
 CC -!- FUNCTION: Negative regulator of ftsZ ring formation; modulates the  
 CC frequency and position of ftsZ ring formation. Inhibits ftsZ ring  
 CC formation at polar sites. Interacts either with ftsZ or with one  
 CC of its binding partners to promote depolymerization (By  
 CC similarity).  
 CC -!- SUBCELLULAR LOCATION: Membrane-associated. Colocalized with ftsZ to  
 CC the nascent septal site (By similarity).  
 CC -!- SIMILARITY: Belongs to the ezrA family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to license@isp-sib.ch).  
 CC -----  
 DR EMBL; AE010101; AAL97460.1; -.  
 DR HAMAP; MF\_00728; -; 1.  
 KW Cell division; Septation; Transmembrane; Coiled coil;  
 KW Complete proteome.  
 FT DOMAIN 1 7 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 8 26 POTENTIAL.  
 FT DOMAIN 27 574 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 102 148 COILED COIL (POTENTIAL).  
 FT DOMAIN 276 366 COILED COIL (POTENTIAL);  
 SQ SEQUENCE 574 AA; 66061 MW; 76A93C1FCABOBRE CRC64;

Query Match 5.8%; Score 101; DB 1; Length 574;  
 Best Local Similarity 21.1%; Pred. No. 3; 2; Mismatches 110; Indels 66; Gaps 12;  
 Matches 60; Conservative 48; MisMatches 110; Indels 66; Gaps 12;

QY 120 TGLEMLINCSKMLPADITQLNNTIPQTESYYDP---NLPPVTKSLVTNCKP-WTDK 172  
 QY :|: ;|: ;|: ;|: ;|: ;|: ;|: ;|: ;|: ;|: ;|: ;|: ;|: ;|:  
 QY 240 TGYERL-----EENYHPPKNEEARFOETIRBSISTRANSELVTL 279  
 QY 173 IHKAYDKDKYR----PEINGEEETAFKMIRTNSHVGOLDDIRK----- 215  
 QY 280 LDRAREENTHIQRIDSLYEVFERELAYKVAKRNQMLPRLAHYKRNNNEQLKDIARL 339  
 QY 216 PRKFVCLNDNIDINHKDQTVKAVALDYE--SMFPIPSQEL 256  
 QY ||: ;|: ;|: ;|: ;|: ;|: ;|: ;|: ;|: ;|: ;|:  
 QY 340 SRKVYL-----SETESLTIVKAPEKOKIETEDSTLVAEQQGL 376  
 QY RESULT 6  
 CENE\_HUMAN STANDARD; PRT; 2663 AA.  
 ID CENE\_HUMAN AC  
 ID Q02224; AC  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Centromeric protein E (CENP-E protein).  
 DE CENE\_E.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN NCBI\_TaxID=866;  
 RN [1]  
 RP SEQUENCE FROM N.A. MEDLINE=95196755; PubMed=1406971;  
 RX RA Thivron D.A., Jordan M.A., Schaar B.T., Szilak I., Cleveland D.W.;  
 RA Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;  
 RT "CENP-E is a putative kinetochore motor that accumulates just before  
 RT mitosis";  
 RL Nature 359:536-539(1992).  
 RN [2]  
 RP CHARACTERIZATION.  
 RX MEDLINE=95196755; PubMed=7889940;  
 RA Chan G.K.T., Schaar B.T., Yen T.J., Wilson J.;  
 RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed  
 RT microtubule motor.;"  
 RL EMBO J. 14:918-926(1995).  
 RN [3]  
 RP CHARACTERIZATION.  
 RX MEDLINE=98437347; PubMed=9763420;  
 RA Chan G.K.T., Schaar B.T., Yen T.J.;  
 RT "Characterization of the kinetochore proteins CENP-E and HBUB1.;"  
 RT interactions with the kintochore proteins CENP-F and HBUB1.;"  
 RL J. Cell Biol. 143:49-63(1998).  
 RN [4]  
 RP FARNESEYLATION.  
 RX MEDLINE=20459117; PubMed=10852915;  
 RA Asrar H.R., James L., Gray K., Carr D., Black S., Armstrong L.,  
 RA Bishop W.R., Kirchmeyer P.;  
 RT "Farnesylyl transferase inhibitors block the farnesylation of CENP-E  
 RT and CENP-F and alter the association of CENP-E with the  
 RT microtubules.;"  
 RL J. Biol. Chem. 275:30451-30457(2000).  
 CC -!- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE  
 CC KINETOCORDE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE  
 CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT  
 CC AND/OR SPINDLE ELONGATION.  
 CC -!- SUBUNIT: INTERACTS WITH CENP-F AND BUB1 KINASE.  
 CC -!- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCORES DURING  
 CC CONGREGATION. RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS  
 CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.  
 CC -!- SIMILARITY: Belongs to the kinesin-like protein family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC EMBL; 215005; CAA78727.1; -.

CC PIR; S28261; S28261.

CC HSSP; PT119; 3KAR.

CC Genew; HGNC1856; CENPE.

DR GK; Q02244; -.

DR MIN; 117143; -.

DR GO; GO:0005334; C:nucleus; TAS.

DR GO; GO:0008350; F:kinetochoore motor activity; TAS.

DR GO; GO:0000667; P:DNA replication and chromosome movement; TAS.

DR GO; GO:0070791; P:miotic chromosome plate congression; TAS.

DR InterPro; IPR001752; Kinesin\_motor.

DR PRINTS; PF00225; KINESIN\_1.

DR SMART; SM0129; KISC\_1.

DR PROSITE; PS00411; KINESIN\_MOTOR\_DOMAIN1; 1.

DR Motor Protein; Cell division; ATP-binding; Coiled coil; Mitosis; Cell cycle; Centromere; Lipoprotein; Preyivation.

FT DOMAIN 336 2471 COILED COIL (POTENTIAL).

FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).

FT NP\_BIND 86 93 ATP (BY SIMILARITY).

FT LIPID 2660 2660 S\_farnesylyl\_Cysteine.

FT SUBSEQUENCE 2663 AA; 312087 MW; CEFPC13880C8C8CB8 CRC64;

Query Match Best Local Similarity 5.8%; Score 101; DB 1; Length 2663; Matches 60; Conservative 47; Mismatches 127; Indels 58; Gaps 10;

QY 33 IDRIVMQLQDMPFEDKTSFHKVRSIEDKOPAFSYFYIIMSAVQPLNISQVFDVTD 92

Db 1972 IQLIQKKEQLQLQYRKYEDVNMHSKINNEQQLQKFEPFNLYKCEMDNFQLTKKH--- 2027

QY 93 QSGVLSIREIRTALTRHHELPISLQDTGLEMLINCSKMPADITQLNNTPPTOSSYYD 152

Db 2028 ---SLEERIVAKERDELRIKESLXMERDQFIAIREMLARDRQHVKPEKELLS 2082

QY 153 PNLLPPPVTKSLVNLCKPKVDKIKHAYKOKNY---RFEIGEEBBAFKMIRTYSHWVG 207

Db 2083 -GQCHLMSLREKSKRKL-KRYSEMDDHYECLMRSLIDEKELEFHRIIMKKLVLS 2140

QY 208 QLDIQRKPRKFWCLND-----NPDHNAHDQT----- 235

Db 2141 YVRKIKBQEHE-CINKFEMDFIDEVERQKELLIKIQLHQDPCDVSPRELDKLNQNMD 2198

QY 236 --VKAVLRFYFESMP-IPSQELPRLPYRNRLPHMELQENRAYR---DXKL 281

Db 2199 LHTEELIKDFSESEPSIKTEFQ-Q-QVLSNRKEMTQFLEELNTRDIEKUK 2248

RESULT 7 TATR\_NPVB  
ID TATR\_NPVB STANDARD; PRT; 587 AA.

AC P33245;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)

DE Trans-activating transcriptional regulatory protein (immediate early DE protein 1) (IE-1).

DE protein 1) (IE-1).

DE IEL1.

GN Bombyx mori nuclear polyhedrosis virus (BmNPV);

OS Bombyx mori nuclear polyhedrosis virus, no RNA stage; Baculoviridae;

OC Nucleopolyhedrovirus.

OX NCBI\_TaxId=10458;

RN [1] SEQUENCE FROM N.A.

OS Viruses, deDNA viruses, no RNA stage; Baculoviridae;

OC Nucleopolyhedrovirus.

OX NCBI\_TaxId=92162753; PubMed=1536895;

OC Huymbrechts R., Guarino L., Van Brussel M., Vulfsteke V.; MIDDLETON; 32:1227-1234(1996).

RX "Nucleotide sequence of a transactivating Bombyx mori nuclear polyhedrosis virus immediate early gene.";

RT polyhedrosis virus immediate early gene.";

RT Biochim. Biophys. Acta 1129:1328-1330(1992).

CC -!- FUNCTION: Regulatory transcriptional protein, which trans-activates gene expression from early baculovirus promoters. Can also trans-activate its own promoter, suggesting that it is auto-regulated during normal infection of insect cells.

CC -!- SIMILARITY: TO OTHER BACULOVIRUSES IEI.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

CC DR PRINS; PR00380; KINESINHEAVY.

CC DR SMART; SM0129; KISC\_1.

CC DR PROSITE; PS00411; KINESIN\_MOTOR\_DOMAIN1; 1.

CC DR Motor Protein; Cell division; ATP-binding; Coiled coil; Mitosis; Cell cycle; Centromere; Lipoprotein; Preyivation.

CC DR DOMAIN 336 2471 COILED COIL (POTENTIAL).

CC DR DOMAIN 2472 2663 GLOBULAR (POTENTIAL).

CC DR NP\_BIND 86 93 ATP (BY SIMILARITY).

CC DR LIPID 2660 2660 S\_farnesylyl\_Cysteine.

CC DR SUBSEQUENCE 2663 AA; 312087 MW; CEFPC13880C8C8CB8 CRC64;

Query Match Best Local Similarity 5.7%; Score 99.5; DB 1; Length 587; Matches 58; Conservative 34; Mismatches 103; Indels 53; Gaps 11;

QY 39 QELOOMFPES----FDTFSHKVRSIEDKOPAFSYFYIIMSAVQPLNISQVFDVTD 92

Db 189 QEITIVFTDFAPVLMRFFDDNDNSNRSDHNSBTGYMFVVKSEVPEPIFAKVSN 248

QY 85 VPEDEVTDQSGVLSDREIRTALRH-----ELPLSPLQDTGLHMLINCSK 131

Db 249 VVYVY-TNNYYMMVNDRVVFVTFDKIRFMISVNLVKGEGIEIPS-QDVNDETAQNCRK 306

QY 132 MFLADITQLANNIPTQESYDPILPVPVTKSLT-----NCKPVTDKXHKAQKDNKY 183

Db 307 CHFVTDH--HTFKALTSYFNLDMYQACTVTIQLSGERKGCGFLSKVYMDQKLF 364

QY 184 RPEIHM-----GEEIATKK-MRTNVHVVGGQDDIRK-----NPRKFCVLDNDIH 229

Db 365 TLPIMLSRKESNEETASNFNSPFSPYVSQLXKSESRKVKPDPNPKNVYVNLIVN 424

QY 230 HKDAOTVK 237

Db 425 KKSILTYK 432

RESULT 8 TATR\_NPVB  
ID TATR\_NPVB STANDARD; PRT; 385 AA.

AC P91153;  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Guanine nucleotide-binding protein alpha-2 subunit (GP-alpha-2).

GN GNA2 OR Ga2.

OS Glycine max (Soybean).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyt; Magnoliophyta; Eudicots; Rosids; OC euerosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine; OX NCBI\_TaxId=847;

RN [1] SEQUENCE FROM N.A.

RC STRAIN=ev; Corsoy; MEDLINE=9715166; PubMed=9002626;

RC Gotor C., Lam E., Cerdido F.J., Romero L.C.; RT isolation and analysis of the soybean Sb2 gene (cDNA), encoding a new member of the plant G-protein family of signal transducers.";

RT Plant Mol. Biol. 32:1227-1234(1996).

CC -!- FUNCTION: Guanine nucleotide-binding proteins (G proteins) are involved as modulators or transducers in various transmembrane signaling systems.

CC -!- SUBUNIT: G proteins are composed of 3 units; alpha, beta and  
 CC gamma. The alpha chain contains the guanine nucleotide binding  
 CC site.  
 CC -!- SIMILARITY: Belongs to the G-alpha family. Subfamily 2 (G(s)).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to license@isb-sib.ch).  
 CC EMBL; X95582; CAAG64834.1; -.  
 DR InterPro; IPR01019; Gprotein\_alpha.  
 DR Pfam; PF00503; G\_alpha; 1.  
 DR ProDom; PDD00281; Gprotein\_alpha; 1.  
 DR SMART; SM0075; G\_alpha; 1.  
 DR GTP-binding; Transducer; ADP-ribosylation.  
 FT NP\_BIND 47 54 GTP (BY SIMILARITY).  
 FT NP\_BIND 220 224 GTP (BY SIMILARITY).  
 FT NP\_BIND 289 292 GTP (BY SIMILARITY).  
 FT MOD\_RES 192 192 ADP-RIBOSYL [1] (BY ACTION OF CTX).  
 SQ SEQUENCE 385 AA; 45096 MW; 7721B1E307067075 CRC64;  
 Query Match 5.7%; Score 99; DB 1; Length 385;  
 Best Local Similarity 22.6%; Pred. No. 44; Mismatches 101; Indels 50; Gaps 11;  
 Matches 57; Conservative 57;  
 Ov 37 VMOBLQDGPPEBDKTT--SFHKYRHEBDMQLAFSVEYIYMSAVQVNI---SOVFDDEV 89  
 DR 56 IFKQIKLFLQTGDEAKLNSYLEVH----AVIQIKLHDGSREFAQN 101  
 DR 90 DITQSGVMSDRERTLATRHLPLSQLDTGLEHMNLNCQMLPADTQLANNIPTQS 149  
 DR 102 DVDSKKVYISNENKEIGKLLGGRL-DYPYL----SKELABEINWKAQET 153  
 DR QY 150 Y---YDPFLPPVTKSLVINKCPTDKHAKYDKNY-KFEMGEERIAKRMIRVN-VSH 204  
 DR 154 YARGSELQDIPDCDYFMENTLQLSDAMVPTEDDVLYARVRTGIVVEIOPSPVGENKKSD 213  
 DR 205 VVGGLDDI---RKNPKEVCLNDNIDHNAQTVKAWLRDPPYESMPPIPSQFELPREYR 261  
 DR 214 EYRLFDGGQRERKRNKHIFEGY---SAVFCIAISESDQTF----EDENR 260  
 DR QY 262 NRFLHHMHSQEW 273  
 DR 261 NRUMETKELPEW 272  
 RESULT 9  
 MYH9 RAT STANDARD PRT; 1961 AA.  
 ID MYH9 RAT  
 AC Q62812;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Myosin heavy chain, nonmuscle type A (Cellular myosin heavy chain,  
 DE type A) (Nonmuscle myosin heavy chain-A) (NMHC-A).  
 GN MYH9.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathii; Muridae; Murinae; Rattus.  
 OC NCBI\_TaxID=1016;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Choi O.H., Park C., Itoh K., Adelstein R.S., Beaven M.A.;  
 RL Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.  
 CC -!- FUNCTION: Cellular myosin appears to play a role in cytokinesis,  
 CC cell shape, and specialized functions such as secretion and  
 CC capping.

CC -!- SUBUNIT: Myosin is a hexameric protein that consists of 2 heavy  
 CC chain subunits (MHC), 2 alkali light chain subunits (MLC) and 2  
 CC regulatory light chain subunits (MLC-2).  
 CC -!- DOMAIN: The rodlike tail sequence is highly repetitive, showing  
 CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,  
 CC characteristic for alpha-helical coiled coils.  
 CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to license@isb-sib.ch).  
 CC EMBL; U11463; AAH74950.1; -.  
 DR HSSP; P10587; IBR2.  
 DR InterPro; IPR00048; IQ\_region.  
 DR InterPro; IPR001609; myosin\_head.  
 DR InterPro; IPR004009; Myosin\_N.  
 DR InterPro; IPR00228; Myosin\_tail.  
 DR Pfam; PF00612; IQ; 1.  
 DR Pfam; PF00063; myosin\_head; 1.  
 DR Pfam; PF02736; Myosin\_N; 1.  
 DR Pfam; PF01576; Myosin\_tail; 1.  
 DR PROSITE; PS00193; MYOSIN\_HEAVY.  
 DR SMART; SM0015; IQ; 1.  
 DR SMART; SM0242; MISC; 1.  
 DR PROSITE; PS00096; IQ; 1.  
 DR Myosin; ATP-binding; Calmodulin-binding; Actin-binding;  
 KW Coiled coil; Multigene family. MYOSIN HEAD-LIKE.  
 FT DOMAIN 779 808 IO.  
 FT DOMAIN 841 1927 COILED COIL (POTENTIAL).  
 FT DOMAIN 174 181 ATP (POTENTIAL).  
 FT DOMAIN 654 676 ACTIN-BINDING.  
 SQ SEQUENCE 1961 AA; 226336 MW; 9887695681FB19E CRC64;  
 Query Match 5.6%; Score 97; DB 1; Length 1961;  
 Best Local Similarity 28.2%; Pred. No. 29;  
 Matches 35; Conservative 26; Mismatches 51; Indels 12; Gaps 5;  
 Ov 102 IRT-LATRHELDSLQDITGLFHMLNCSQMLPADTOL-NNIPQPSVSYDPNNPPVT 159  
 DR 1267 VRTBLADKVSKQVLDSDVUTGLNQSDSKSKUTQFSALESQLODQBLQENQKL- 1325  
 DR QY 160 KSLVTKCPVDKIHAKYDKNYKFEMGEERIAKRMIRVN-VSHVUGQOLDIKPRKF 219  
 DR 1326 -SLSRKLIQMED----EKNSPREQLEEEBREAKNLKQIAITLIAQVTDWKKNEDG 1377  
 DR 220 V-CL 222  
 DR 1378 VGCL 1381  
 RESULT 10  
 VAC1 YEAST STANDARD PRT; 515 AA.  
 ID VAC1 YEAST  
 AC P32609;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE VAC1 protein.  
 GN VAC1 OR PEP7 OR VPS19 OR YDR323C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes.  
 OC Saccaromycetales; Saccaromycetaceae; Saccharomyces.  
 OC NCB\_ TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=LWY148; PDBP2\_PLAVB  
 RX MEDLINE=92112720; PubMed=1730622;  
 RA Weisman L.S.; Wickner W.;  
 RT "Molecular characterization of VAC1, a gene required for vacuole  
 inheritance and vacuole protein sorting";  
 RL J. Biol. Chem. 267:618-623(1992).  
 CC -!- FUNCTION: REQUIRED FOR VACUOLE SEGREGATION AND VACUOLE PROTEIN  
 SORTING. POSSIBLY PART OF A COMPLEX WHICH TETHERS THE VACUOLE  
 MEMBRANE TO MICROTUBULES, EITHER DIRECTLY OR VIA KINESIN OR  
 DYNENIN-LIKE MOTOR PROTEINS. PROBABLY FUNCTIONS IN SEVERAL  
 INTERORGANELLE TRAFFIC PATHWAYS.  
 CC -!- SUBCELLULAR LOCATION: PERIPHERALLY BOUND ELEMENT OF THE  
 CYTOSKELETON OR VACUOLE (POSSIBLY)  
 CC -!- SIMILARITY: Contains 1 C2H2-type zinc finger.  
 CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-109 IS THE  
 INITIATOR.

This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to license@ib-sib.ch).

CC EMBL; M80596; AAA35203.1; -.  
 DR Germoline; 140315; -.  
 DR SGD; S000731; PEP7.  
 DR GO; GO:005737; C:cytoplasm; IDA.  
 DR GO; GO:019897; C:extrinsic to plasma membrane; IDA.  
 DR GO; GO:005904; R:nonspecific vesicle docking; IMP.  
 DR GO; GO:000906; Znf\_C2H2.  
 DR InterPro; IPR000306; Znf\_FYVE.  
 DR Pfam; PF01363; FIVE; 1.  
 DR Pfam; PF00096; zf-C2H2; 1.  
 DR SMART; SM0064; FYVE; 2.  
 DR SMART; SM0335; Znf\_C2H2; 1.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 1.  
 DR KW Zinc-finger; Repeat.  
 FT ZN\_FING 6 29 C2H2-TYPE.  
 FT ZN\_FING 72 137 FYVE-TYPE 1.  
 FT ZN\_FING 215 297 FYVE-TYPE 2.  
 SQ SEQUENCE 515 AA; 59469 MW; 02C204EBCA8CACQ CRC64;

Query Match 5.5%; Score 96; DB 1; Length 515;  
 Best Local Similarity 22.4%; Pred. No. 6.6; Mismatches 96; Indels 62; Gaps 12;  
 Matches 57; Conservative 39; Mismatches 96; Indels 62; Gaps 12;

OY 28 HMP-----HMDRIVQBLDMFFBFEDKTSFHKVRSDFMOPAFSYVYLMASVQI 80  
 OY :|:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:  
 Db 281 HIPISIRLCSHCIDML-----FGRKENK---DVRMLPLSGPAKDSQ--- 321

OY 81 NTSQVFEDVDTOSGVISDRERTLAIRHELI-PUSLQDTGLEMLNSKMLPADIQ 139  
 OY |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:  
 Db 322 NISKVDSLIPFEDSLSNLKETAKDSEVNLDPKNUNLJARLYCYLSFNLTILQ 381

OY 140 L-----NNIPPTQESYDNPPLPVTKSLVTCKPVTDKTHAYKDKNKFYR 185  
 OY :|:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:  
 Db 382 LISVERPQSHLERQLONSIKIASAYINEKILP-KSLPATLNP--EGHKTNEGQKAEP 437

OY 186 EI----MGEEIAFK-----MRTYVSHWV-GOLDIIRKPF--KFCVGLNDNTFH 228  
 OY :|:|||:|||:|||:|||:|||:|||:|||:|||:|||:  
 Db 438 EVKULSOLMILENTIREKVELREELMVKEOSYLLISTODYKKORLLEEVINKNL 497

OY 229 NHKDAQTVKAVRD 242  
 OY :|:|||:|||:|||:  
 Db 498 LHSRIHTVQSKUGD 511

RESULT 11  
 RBP2\_PLAVB ID\_RBP2\_PLAVB STANDARD; PRT; 2867 AA.  
 AC Q00799; Q9N2M3;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 15-MAR-2004 (Rel. 43, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Reticulocyte binding protein 2 precursor (PvRBP-2).  
 RN RBP-2 OR RBP2.  
 OS Plasmodium vivax (strain Belem).  
 OC Buryatoya; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=31273;  
 RN [1] RP SEQUENCE FROM N.A., AND REVISIONS TO 2438-2439.  
 RA RBP-2; Published=0838229;  
 RT Plasmodium vivax reticulocyte binding protein-2 (PvRBP-2) shares  
 structural features with PvRBP-1 and the Plasmodium yoelii 235 kDa  
 rhabtry protein family.";  
 RL Mol. Biochem. Parasitol. 108:257-262(2000).  
 RN [2] RP SEQUENCE OF 1189-2439 FROM N.A.  
 RA MEDLINE=9231538; PubMed=161771;  
 RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;  
 RT "A reticulocyte-binding protein complex of Plasmodium vivax  
 merozoites".  
 RL Cell 69:1213-1226(1992).  
 CC -!- FUNCTION: Involved in reticulocyte adhesion. Specifically binds to  
 CC human reticulocyte cells.  
 CC -!- SUBCELLULAR LOCATION: Membrane-bound (Probable).  

This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to license@ib-sib.ch).

CC EMBL; AR184633; ARF76523.1; -.  
 DR DR HSSP; P03069; IGCN.  
 DR KW Malaria; Receptor; Signal; Transmembrane; Repeat.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT CHAIN 22 2867 RETICULOCYTE BINDING PROTEIN 2.  
 FT DOMAIN 22 2805 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 2805 2867 POTENTIAL.  
 FT DOMAIN 44 133 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 560 758 ASN-RICH.  
 FT DOMAIN 1112 1285 LYS-RICH.  
 FT DOMAIN 2758 2785 LYS-RICH.  
 FT REPEAT 2758 2761 7 X 4 AA TANDEM REPEATS OF H-D-D-T.  
 FT REPEAT 2762 2765 1.  
 FT REPEAT 2766 2769 2.  
 FT REPEAT 2770 2773 3.  
 FT REPEAT 2774 2777 4.  
 FT REPEAT 2778 2781 5.  
 FT REPEAT 2782 2782 6.  
 SQ SEQUENCE 2867 AA; 331433 MW; 6ED8CA71AFBF0D3 CRC64;

Query Match 5.5%; Score 95; DB 1; Length 2867;  
 Best Local Similarity 20.5%; Pred. No. 65; Mismatches 78; Indels 84; Gaps 14;  
 Matches 56; Conservative 55; Mismatches 78; Indels 84; Gaps 14;

OY 7 LRVYK-----ILNSKFGFNSRKPKAHMHPMIDRIVQWQLQDWPPEERDKTSFHKV 58  
 OY :|:|||:|||:|||:|||:|||:|||:|||:|||:|||:  
 Db 2193 LAVYKRNEDTVQDVLTNEHENRNTKQSNH-----PTNFDSN--- 2232

OY 59 HSEBLQFQFSYVYLMASVQPNFED-VDTOSGVLSDRERIRTLATRHELPSL 116  
 OY :|:|||:|||:|||:|||:|||:|||:  
 Db 2233 SSEBLQFQFSYVYLMASVQPNFED-VDTOSGVLSDRERIRTLATRHELPSL 2288  
 OY 117 QDL-----TGLEHMLNCSKMLPADITQ-NNIPPIQESTYDPNLPPTVSLVNTCKPV 169

RESULT 12

ID	2299 NEYQTSNEVKGQEMKSNADKV--DYSKIFIVLQKS-----NIVNQHSI 2335	FT ACT-SITE 422	422	POTENTIAL.
AC	P15274; 1990 (Rel. 14, Created)	FT ACT-SITE 631	631	POTENTIAL.
DT	01-FEB-1996 (Rel. 33, Last sequence update)	FT ACT-SITE 707	707	POTENTIAL.
DE	AMP deaminase (Rel. 42, Last annotation update)	FT ACT-SITE 708	708	POTENTIAL.
GN	AMD1 OR AMD OR YNL035C.	FT CONFLICT 568	568	F -> C (IN REF. 1).
OC	Saccharomyces cerevisiae (Baker's yeast).	FT SEQUENCE 810 AA;	93301 MW;	7A6DCB43B9B45C93 CRC64;
OX	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.			
RN	[1] NCBI_TaxID:4932;			
RP	SEQUENCE FROM N.A.			
RK	MEDLINE=9015403; PubMed=9169872;			
RA	Meyer S.L., Kvalnes-Krick K.L., Schramm V.L.; "Characterization of AMD, the AMP Deaminase gene in yeast. Production of and strain, cloning, nucleotide sequence, and properties of the protein.", Biochemistry 28: 8734-8743(1989).			
RA	Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T., Connor R., Dedman K., Devin S., Hamlin N., Hunt S., Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A., Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.; "the nucleotide sequence of <i>Saccharomyces cerevisiae</i> chromosome XII.", Nature 387:90-93(1997).			
RL	Nature 387:90-93(1997);			
CC	-!- FUNCTION: AMP deaminase plays a critical role in energy metabolism.			
CC	-!- CATALYTIC ACTIVITY: AMP + H <sub>2</sub> O = IMP + NH <sub>3</sub> .			
CC	-!- PATHWAY: Purine nucleotide cycle.			
CC	-!- SUBUNIT: Homotetramer.			
CC	-!- SIMILARITY: Belongs to the adenosine and AMP deaminases family.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@isb-sib.ch).			
CC	SGD; SGD-0004498; AMD1.			
DR	GO:0003876; P:AMP deaminase activity; IDA.			
DR	GO:0006163; P:purine nucleotide metabolism; IC.			
DR	InterPro; IPR004650; A/AMP deaminase.			
DR	InterPro; IPR001365; A/AMP deaminase.			
DR	InterPro; IPR00329; AMP_deaminase.			
DR	Pram; PF00362; A/deaminase; 1.			
DR	TIGR04129; AMP deaminase; 1.			
DR	PROSITE; PS00485; A_DRAMINASE; 1.			
KW	Hydroxylase; Nucleotide metabolism.			

RESULT 13

ID	20 FTSRKVPAHMP--MIDRIVQELQDMPPERFDKSFHKVRHSEBMQFASFYVILMSA 76	FT ACT-SITE 422	422	POTENTIAL.
AC	354 UNP NR_0012777; Conservative 51; Mismatches 114; Indels 152; Gaps 19; Matches 77;	FT ACT-SITE 631	631	POTENTIAL.
DT	01-NOV-1995 (Rel. 32, Last sequence update)	FT ACT-SITE 707	707	POTENTIAL.
DE	01-NOV-1995 (Rel. 32, Last annotation update)	FT ACT-SITE 708	708	POTENTIAL.
GN	YAL19W OR G73594.	FT CONFLICT 568	568	F -> C (IN REF. 1).
OC	<i>Saccharomyces cerevisiae</i> (Baker's yeast).	FT SEQUENCE 810 AA;	93301 MW;	7A6DCB43B9B45C93 CRC64;
OX	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.			
RN	[1] NCBI_TaxID:4932;			
RP	SEQUENCE FROM N.A.			
RK	MEDLINE=9015403; PubMed=9169872;			
RA	Meyer S.L., Kvalnes-Krick K.L., Schramm V.L.; "Characterization of AMD, the AMP Deaminase gene in yeast. Production of and strain, cloning, nucleotide sequence, and properties of the protein.", Biochemistry 28: 8734-8743(1989).			
RA	Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T., Connor R., Dedman K., Devin S., Hamlin N., Hunt S., Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A., Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.; "the nucleotide sequence of <i>Saccharomyces cerevisiae</i> chromosome XII.", Nature 387:90-93(1997).			
RL	Nature 387:90-93(1997);			
CC	-!- FUNCTION: AMP deaminase plays a critical role in energy metabolism.			
CC	-!- CATALYTIC ACTIVITY: AMP + H <sub>2</sub> O = IMP + NH <sub>3</sub> .			
CC	-!- PATHWAY: Purine nucleotide cycle.			
CC	-!- SUBUNIT: Homotetramer.			
CC	-!- SIMILARITY: Belongs to the adenosine and AMP deaminases family.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@isb-sib.ch).			
CC	SGD; SGD-0004498; AMD1.			
DR	GO:0003876; P:AMP deaminase activity; IDA.			
DR	GO:0006163; P:purine nucleotide metabolism; IC.			
DR	InterPro; IPR004650; A/AMP deaminase.			
DR	InterPro; IPR001365; A/AMP deaminase.			
DR	InterPro; IPR00329; AMP_deaminase.			
DR	Pram; PF00362; A/deaminase; 1.			
DR	TIGR04129; AMP deaminase; 1.			
KW	Hydroxylase; Nucleotide metabolism.			

RESULT 14

ID	20 FTSRKVPAHMP--MIDRIVQELQDMPPERFDKSFHKVRHSEBMQFASFYVILMSA 76	FT ACT-SITE 422	422	POTENTIAL.
AC	354 UNP NR_0012777; Conservative 51; Mismatches 114; Indels 152; Gaps 19; Matches 77;	FT ACT-SITE 631	631	POTENTIAL.
DT	01-NOV-1995 (Rel. 32, Last sequence update)	FT ACT-SITE 707	707	POTENTIAL.
DE	01-OCT-1996 (Rel. 34, Last annotation update)	FT ACT-SITE 708	708	POTENTIAL.
GN	YAL19W OR G73594.	FT CONFLICT 568	568	F -> C (IN REF. 1).
OC	<i>Saccharomyces cerevisiae</i> (Baker's yeast).	FT SEQUENCE 810 AA;	93301 MW;	7A6DCB43B9B45C93 CRC64;
OX	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.			
RN	[1] NCBI_TaxID:4932;			
RP	SEQUENCE FROM N.A.			
RK	MEDLINE=9015403; PubMed=9169872;			
RA	Guerreiro P., Mala e Silva A., Barreiros T., Arrovo J., Garcia-Gonzalez M., Garcia-Saez M.I., Rodrigues-Pousada C., Nonato C.; "The complete sequence of a 9000 bp fragment of the right arm of <i>Saccharomyces cerevisiae</i> chromosome VII contains four previously unknown open reading frames.", Yeast 11:1087-1091(1995).			
RA	Guerrero P., Mala e Silva A., Barreiros T., Arrovo J., Garcia-Gonzalez M., Garcia-Saez M.I., Rodrigues-Pousada C., Nonato C.; "The complete sequence of a 9000 bp fragment of the right arm of <i>Saccharomyces cerevisiae</i> chromosome VII contains four previously unknown open reading frames.", Yeast 11:1087-1091(1995).			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@isb-sib.ch).			
CC	SGD; SGD-0004498; AMD1.			
DR	GO:0003876; P:AMP deaminase activity; IDA.			
DR	GO:0006163; P:purine nucleotide metabolism; IC.			
DR	InterPro; IPR004650; A/AMP deaminase.			
DR	InterPro; IPR001365; A/AMP deaminase.			
DR	InterPro; IPR00329; AMP_deaminase.			
DR	Pram; PF00362; A/deaminase; 1.			
DR	TIGR04129; AMP deaminase; 1.			
KW	Hydroxylase; Nucleotide metabolism.			



FT	HELIK	80	82
FT	TURN	83	84
FT	STRAND	87	87
FT	TURN	88	89
FT	HELIK	96	108
FT	TURN	109	110
FT	STRAND	113	115
FT	STRAND	120	123
FT	TURN	130	131
FT	STRAND	134	139
FT	TURN	140	142
FT	HELIK	145	147
FT	HELIK	152	166
FT	TURN	167	167
FT	HELIK	170	175
FT	TURN	178	179
FT	HELIK	182	197
FT	HELIK	218	223
FT	TURN	224	225
FT	HELIK	226	234
FT	STRAND	235	235
FT	TURN	242	243
FT	STRAND	244	245
FT	STRAND	248	255
FT	TURN	257	258
FT	STRAND	261	269
FT	HELIK	273	276
FT	TURN	277	277
FT	STRAND	281	282
FT	HELIK	287	295
FT	TURN	298	301
FT	HELIK	302	305
FT	TURN	309	311
FT	TURN	313	314
FT	HELIK	316	317
FT	TURN	327	341
FT	HELIK	342	342
FT	TURN	345	363
FT	STRAND	367	367
FT	STRAND	376	376
FT	HELIK	381	390
FT	TURN	391	391
FT	HELIK	394	402
FT	TURN	419	449
FT	STRAND	458	464
FT	HELIK	474	474
FT	TURN	505	505
FT	HELIK	507	508
FT	HELIK	521	528
FT	HELIK	536	543
FT	HELIK	551	561
FT	TURN	562	564
FT	TURN	566	567
FT	STRAND	568	570
FT	STRAND	579	584
FT	TURN	585	586
FT	STRAND	587	592
FT	TURN	594	595
FT	HELIK	596	601
FT	HELIK	606	613
FT	TURN	614	614
FT	HELIK	618	623
FT	TURN	624	624
FT	HELIK	658	674
FT	TURN	675	675
FT	STRAND	677	684
FT	TURN	692	693
FT	HELIK	697	706
FT	TURN	707	708
FT	HELIK	718	725
FT	STRAND	726	733

RESULT 15  
SYN\_BORBU ID SYN\_BORBU STANDARD; PRT; 353 AA.  
AC 0510538;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE TRYPTOPHANYL-tRNA synthetase (EC 6.1.1.2) (Tryptophan-tRNA ligase (TTRS)).  
DE TRRS OR TRSA OR BB005.  
OS Borrelia burgdorferi (Lyme disease spirochete).  
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.  
OX NCBI\_TaxID:133;  
RN [1] SEQUENCE FROM N.A.  
RP STRAIN=ATCC 35210 / B31;  
RC MEDLINE:98065943; PubMed:9403685;  
RA Frazer C.M., Castens S., Huang W.M., Sutton G.G., Clayton R.A.,  
RA Lathigra R., White O., Kerchuk K.A., Dodson R., Hickey E.K., Gwinn M.,  
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,  
RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,  
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,  
RA Utterback T., Wathey L., McDonald L., Artiach P., Bowman C.,  
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,  
RA Smith H.O., Venter J.C.;  
RT "Genomic sequence of a Lyme disease spirochete, Borrelia  
burgdorferi.";  
RL Nature 390:880-886(1997).  
-- CATALYTIC ACTIVITY: ATP + L-tryptophan + tRNA(Trp) = AMP +  
-- diphosphate + L-tryptophanyl-tRNA(Trp).  
-- SUBUNIT: Homodimer (By similarity).  
-- SUBCELLULAR LOCATION: Cytoplasmic.  
-- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.  
-- CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
-- between the Swiss Institute of Bioinformatics and the EMBL collaboration -  
-- the European Bioinformatics Institute. There are no restrictions on its  
-- use by non-profit institutions as long as its content is in no way  
-- modified and this statement is not removed. Usage by and for commercial  
-- entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
-- or send an email to license@isb-sib.ch).  
-- CC EMLI, AE00115, AAC63398.1; -.  
DR PIR; E70100; E70100.  
DR HSSP; P00953; 1D2R.

DR TIGR; BB0005; -.  
 DR HAMAP; MF\_00140; -; 1.  
 DR InterPro; IPR022305; tRNA-synt\_1b.  
 DR InterPro; IPR01412; tRNA-synt\_1.  
 DR Pfam; PF00579; tRNA-synt\_1b; 1.  
 DR PRINTS; PR01039; TRNA-SYNTTRP.  
 DR TIGRFMS; TIGR00233; trts; 1.  
 DR PROSITE; PS00178; AA-TRNA-LIGASE\_I; 1.  
 KW Aminocyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
 Complete proteome.  
 FT SITE 14 22 "HIGH" REGION.  
 FT SITE 200 204 "KRSKS" REGION.  
 FT BINDING 203 ATP (BY SIMILARITY).  
 SQ SEQUENCE 353 AA; 40419 MW; 3ADD5455899D40BF CRC64;

---

Query Match Similarity Score 93.5; DB 1; Length 353;  
 Best Local Similarity 18.2%; Pred. No. 6.4.; Gaps 15;  
 Matches 56; Conservative 47; Mismatches 91; Indels 113; Gaps 15;

```

QY 28 RMPHMDRIMQ-ELQDMFPPFEDKTSFHKYRHSDM-----QFAFSYF---- 70
Db 19 HLGHYVGSVWLRKFQEETYFIILADLHILITTKFDLKSNTIFSNVRENWLVDIAGG 78
QY 71 -----YTIMSAVQPL-NISQYFDEVDTDOGVLSDREIRLATRHELLLSLCILT--G 121
Db 79 PDKVSIYLQSAIPELFBLHLIFSMIVT-----VARLQRIPIK-SIKUMSIAG 123

```

---

QY 122 LEHM--LINGSKMLPADI-TQLNNIPPOES-----YDPMPLPPVTK 160
 Db 124 LKEIPYGLLGYPVIMSADILMTKANLVPGRDNESHTIEFARARRFHNLKNNFPPIE 183

QY 161 SIVTNCKPVT-----DKHFKAYKDKNKCFPPEINGBEE- 192

Db 184 SVFTDERSPLVGIYGKOKMSKLDNAFLNDNDENLJEKKINSMYDPNRTRADIGVNEV 243
 QY 193 --IAFKMIRTVSHV-----VGQDIDTRKNPRKFVCLNDNIDHNNHKAQTYKAV- 239
 Db 244 PIFIYHSFFNSHEEVDLKSRVYKGKVGVKEVKKLFALAIN-----SELIKPIRD 293

QY 240 LRDFYES 246
 Db 294 KRSFYEA 300

Search completed: July 26, 2004, 11:09:44  
 Job time : 6.46316 secs

This Page Blank (uspto)

Copyright GenCore version 5.1.6  
(c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2004, 11:05:03 ; Search time 21.4324 Seconds  
(without alignments)

4828.666 Million cell updates/sec

Title: US-10-657-280-2

Perfect score: 1731

Sequence: DTFADSLRVNKILNSKFGP.....RKIFPRRIKHEASPNRIRV 328

Scoring table: BLASTM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SPREMBL\_25;\*

1: sp\_archaea;\*

2: sp\_bacteria;\*

3: sp\_fungi;\*

4: sp\_human;\*

5: sp\_invertebrate;\*

6: sp\_mammal;\*

7: sp\_mhc;\*

8: sp\_orcanelle;\*

9: sp\_phage;\*

10: sp\_plant;\*

11: sp\_rabbit;\*

12: sp\_virus;\*

13: sp\_vertebrate;\*

14: sp\_unclassified;\*

15: sp\_rvirus;\*

16: sp\_bacteriap;\*

17: sp\_archeap;\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

**SUMMARIES**

Result No.	Score	Query Match Length	DB ID	Description
1	1731	100.0	490	Q86TQ2
2	1731	100.0	950	4 Q86TQ2
3	1290	74.5	248	4 Q8NPW9
4	704	40.7	132	4 Q8NVA4
5	581	33.6	384	11 Q61340
6	506	29.2	666	5 Q8X14
7	506	29.2	666	5 Q9V533
8	113.5	6.6	545	2 Q86215
9	113.5	6.6	545	2 Q8AD00
10	113.5	6.6	545	2 Q8ACZ9
11	113.5	6.6	545	2 Q83U59
12	113.5	6.6	545	16 Q9JWWB
13	111	6.4	1048	15 Q26023
14	111	6.4	3119	5 Q25857
15	111	6.4	3119	5 Q814T3
16	106.5	6.2	483	5 Q86KMO

**ALIGNMENTS**

17 106 6.1 574 16 Q9AOK9 Q9aq0kg streptococc  
18 106 6.1 739 8 Q9WU79 Q9mj79 physarum po  
19 106 6.1 2148 5 Q9BD3 Q9bjd3 physarum po  
20 105 6.1 666 13 Q90XS4 Q90xs4 xenopus lae  
21 104.5 6.0 1999 11 Q93731 Q93731 rattus norv  
22 104 6.0 424 5 Q8188 Q8i188 plasmidium  
23 104 6.0 574 16 Q9K839 Q9k839 streptococc  
24 103 6.0 621 16 Q913R0 Q913r0 bacillus ce  
25 103 6.0 690 17 Q87915 Q87915 pyrococcus  
26 102.5 5.9 676 5 Q7YX23 Q7yx23 cryptospori  
27 102.5 5.9 706 8 Q98594 Q98594 guillardia  
28 102.5 5.9 515 3 Q98229 Q99229 sachcharomyces  
29 101.5 5.9 301 16 Q8Y9G9 Q99229 listeria mo  
30 101.5 5.9 408 12 Q9J575 Q9j575 fowl-pox vir  
31 101 5.8 375 10 Q7XWH9 Q7xwh9 oryza sativ  
32 101 5.8 574 16 Q8P1M3 Q8p1m3 streptococc  
33 100.5 5.8 265 16 Q8D241 Q8d241 wiggleswort  
34 100.5 5.8 468 5 Q8EBM6 Q8ebm6 drosophila  
35 100 5.8 5987 5 Q8TDN3 Q8tdn3 plasmidium  
36 99.5 5.7 893 4 Q910L3 Q910l3 homo sapien  
37 99 5.7 723 16 Q8RH48 Q8rh48 fusobacteri  
38 98 5.7 5687 5 Q8JUH4 Q8juh4 plasmidium  
39 97.5 5.6 735 16 Q51543 Q51543 borrelia bu  
40 97.5 5.6 893 4 Q81YMO Q81ymo homo sapien  
41 97.5 5.6 1498 16 Q92FP9 Q92fp9 listeria in  
42 97 5.6 516 5 Q18679 Q18679 caenorhabdi  
43 97 5.6 620 16 Q99227 Q99227 streptococc  
44 97 5.6 776 16 Q8B816 Q8b816 shewanella  
45 97 5.6 942 15 Q911I4 Q911i4 streptomyce

Db	283 GLEMMLNSKMPADIDVNLNPPTOSYYPNLPTVKSIVPDKVPLKPKVCLNDNHHQDAQTVKAVL 342	Db	863 RDPYESMFPISQFELPREYRNPFLHMELOEWRAYRDKLKFWMTHCVLATLIMFTFSFF 922
OY	181 NCKRFPEIIGEEETAFKMRITNVSHVGGLDDDRKKRPFKVCLNDNHHQDAQTVKAVL 240	OY	301 AEQLALKRKIFPRRIKEASPNRIRV 328
Db	343 NCKRFPEIIGEEETAFKMRITNVSHVGGLDDDRKKRPFKVCLNDNHHQDAQTVKAVL 402	Db	923 AEQLALKRKIFPRRIKEASPNRIRV 950
OY	241 RDPYESMFPISQFELPREYRNPFLHMELOEWRAYRDKLKFWMTHCVLATLIMFTFSFF 300		
Db	403 RDPYESMFPISQFELPREYRNPFLHMELOEWRAYRDKLKFWMTHCVLATLIMFTFSFF 462		
OY	301 AEQLALKRKIFPRRIKEASPNRIRV 328		
Db	463 AEQLALKRKIFPRRIKEASPNRIRV 490		
RESULT 2			
ID	Q9NLU2 PRELIMINARY; PRT; 950 AA.	ID	Q9NPM9 PRELIMINARY; PRT; 248 AA.
AC	Q9NLU2;	AC	Q9NPM9;
DT	01-MAY-2000 (TREMBlrel. 13, Created)	DT	01-OCT-2000 (TREMBlrel. 15, Created)
DT	01-OCT-2002 (TREMBlrel. 22, Last sequence update)	DT	01-OCT-2002 (TREMBlrel. 15, Last sequence update)
DE	Hypothetical protein KIAA1208 (Fragment).	DE	01-OCT-2000 (TREMBlrel. 25, Last annotation update)
GN	KIAA1208	GN	Hypothetical protein (Fragment).
OS	Homo sapiens (Human).	OS	DKFPP762B226.
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	OC	Homo sapiens (Human).
OC	Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606; [1]		NCBI_TaxID=9606; [1]	
RN	SEQUENCE FROM N.A.	RN	
RC	TISSUE=Brain;	RC	
RA	Medline=20039619; PubMed=10574462;	RA	
RA	"Prediction of the coding sequences of unidentified human genes. XV: The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro.", DNA Res. 6:337-345 (1999)	RA	
RL	EMBL; AB033934; BX486522; -.	RL	Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AB033934; BX486522; -.	DR	EMBL; AL35988; CAB94874.1; -.
DR	GO; GO:0005509: F:calcium ion binding; IEA.	DR	PIR; T50618; T50618.
DR	GO; GO:0030154; P:cell differentiation; IEA.	DR	GO; GO:0005509: F:calcium ion binding; IEA.
DR	InterPro; IPR002448; EF-hand.	DR	InterPro; IPR002048; EF-hand.
DR	InterPro; IPR000010; Notch_dom.	DR	Prosite; PS00018; EF_HAND; 1.
DR	PFAM; PF00036; eFhand_1.	DR	Hypothetical protein.
DR	PFAM; PF00066; notch_1.	FT	NON_TER 1 1
DR	SMART; SM00004; NL_2.	SEQUENCE 248 AA; 29336 MW; D4C4C4A6C4505522 CRC64;	
DR	PROSITE; PS00018; EF_HAND; 1.	Query Match 74.5%; Score 1290; DB 4; Length 248;	
KW	Hypothetical protein.	Best Local Similarity 100.0%; Pred. No. 3.4e-94; Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
FT	SQIQUENCE 950 AA; 108785 MW; F653BDEA6EAC7503 CRC64;	Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Query Match 100.0%; Score 1731; DB 4; Length 950; Best Local Similarity 100.0%; Pred. No. 3.e-128; Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	QY	85 VFDDEVDTDQGSVLSDRETTATRHEPLSLQDLTGLEMLINCSKMLPADITQTNIP 144	
OY	1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60	5 VFDEVDTDQGSVLSDRETTATRHEPLSLQDLTGLEMLINCSKMLPADITQNNIP 64	
QY	2 1 DTFAADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60	QY	145 PTBDSYYPNPLPVTKSIVTNCVKPTVTKHAKYDKKQYREBIMGSEETAKMIRTVSH 204
OY	3 Q9NLU2 PRELIMINARY; PRT; 950 AA; 108785 MW; F653BDEA6EAC7503 CRC64;	Db	65 PTQBSYYDNLPVTKSIVTNCVKPTVTKHAKYDKKQYREBIMGSEETAKMIRTVSH 124
QY	4 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60	QY	205 VFGQDPLDRKNPKVFCVLNDNHHQDAQTVKAVRDYFESMFPISQFELPREYRNP 264
OY	5 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60	Db	125 VFGQDPLDRKNPKVFCVLNDNHHQDAQTVKAVRDYFESMFPISQFELPREYRNP 184
QY	6 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60	QY	265 LHMELQEWRAYRDKLKFWMTHCVLATLIMFTFSFF 324
OY	7 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60	Db	185 LHMELQEWRAYRDKLKFWMTHCVLATLIMFTFSFF 244
QY	8 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60	QY	325 RIRV 328
QY	9 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60	Db	245 RIRV 248
QY	10 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	11 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	12 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	13 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	14 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	15 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	16 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	17 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	18 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	19 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	20 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	21 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	22 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	23 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	24 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	25 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	26 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	27 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	28 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	29 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	30 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	31 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	32 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	33 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	34 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	35 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	36 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	37 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	38 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	39 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	40 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	41 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	42 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	43 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	44 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	45 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	46 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	47 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	48 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	49 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	50 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	51 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	52 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	53 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	54 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	55 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	56 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	57 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	58 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	59 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	60 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	61 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	62 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	63 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	64 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	65 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	66 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	67 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	68 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	69 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	70 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	71 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	72 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	73 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	74 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	75 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	76 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	77 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	78 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	79 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	80 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	81 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	82 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	83 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	84 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	85 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	86 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	87 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	88 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	89 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	90 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	91 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	92 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	93 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	94 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	95 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	96 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	97 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	98 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	99 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	100 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	101 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	102 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	103 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	104 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	105 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	106 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	107 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	108 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	109 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	110 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	111 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	112 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	113 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	114 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	115 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	116 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	117 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	118 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	119 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	120 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	121 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	122 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	123 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	124 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	125 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	126 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	127 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	128 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	129 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	130 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	131 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	132 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
QY	133 1 DTFADSLIYVNLNSKGFTSRKVAPMHPMDRIVYQELQDMFEEFDKTSFKVRS 60		
Q			

RP SEQUENCE FROM N.A.  
 RC TISSUE-Placenta;  
 RA Isogai T., Ota T., Hayashi K., Sugano S., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Togita S., Komai F., Hara R., Takeuchi K.,  
 RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,  
 RA Wakamatsu A., Nakamura Y., Naohari K., Masuho Y., Oshima A.;  
 RT "NEDO human cDNA sequencing project.";  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AK001821; BAA919261.; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 132 AA; 16083 MW; 92B5DEB4E062AD07 CRC64;

Query Match 40.7%; Score 704; DB 4; Length 132;  
 Best Local Similarity 100.0%; Pred. No. 3.6e-48;  
 Matches 132; Conservative 0; MisMatches 0; Indels 0; Gaps 0;

Qy 197 MRTNTSHVGQLDDIRKRNPKRKEVCLNDNDHDKAQTVGAVDIFYESMPPPSQFEL 256  
 Db 1 MIRTNVSHVGQLDDIRKRNPKRKEVCLNDNDHDKAQTVGAVDIFYESMPPPSQFEL 60

Qy 257 PREYRNPLHMELOQENRAKRDKEFICLNLIMFTSPFABQLALKRKPFPR 316  
 Db 61 PREYRNPLHMELOQENRAKRDKEFICLNLIMFTSPFABQLALKRKPFPR 120

Qy 317 IKKEASPNRIRV 328  
 Db 121 IKKEASPNRIRV 132

RESULT 5

ID 061340 PRELIMINARY; PRT; 384 AA.  
 AC 061340:  
 DT 01-NOV-1996 [TREMBrel. 01, Created]  
 DT 01-NOV-1996 [TREMBrel. 01, Last sequence update]  
 DT 01-OCT-2003 [TREMBrel. 25, Last annotation update]  
 DE Basic domain/leucine zipper transcription factor (Fragment).  
 OS MAFB.  
 OC Mus musculus (Mouse)  
 OC Bokarotta; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TAXID:9909;  
 RN [1] SEQUENCE FROM N.A.

RP SEQUENCE FROM N.A.  
 RX MEDLINE=95094266; PubMed=8001130;

RA Cordes S.P., Barsh G.S.;  
 RT "The mouse segmentation gene kr encodes a novel basic domain-leucine zipper transcription factor.";  
 RT Cell 79:1025-104 (1994);  
 RL EMBL; L36434; AAA65688.1; -.  
 DR PIR; 149528; I4528.  
 DR MGD; MGI:104555; Mafb.  
 DR CO; GO:003677; F:DNA binding; IDA.  
 DR GO; GO:0006357; P:regulation of transcription from Pol II pro. . . ; IDA.  
 DR InterPro; IPR00048; EF-hand.  
 DR Pfam; PF00036; eifhand\_1.  
 DR PROSITE; PS00018; EF\_HAND\_1.  
 FT NON\_TER; 1.  
 SQ SEQUENCE 384 AA; 43754 MW; 51F473C8807A7555 CRC64;

Query Match 33.6%; Score 581; DB 11; Length 384;  
 Best Local Similarity 97.4%; Pred. No. 7.3e-38;  
 Matches 112; Conservative 2; MisMatches 1; Indels 0; Gaps 0;

Qy 1 DTFADSLRVNKILNSKGFSRKPVPHMPIDRIVQELQDMPPZEFKTSFHKYRH 60  
 Db 267 DTFADSLRVNKILNSKGFSRKPVPHMPIDRIVQELQDMPPZEFKTSFHKYRH 326

Qy 61 EDMQFASPYYIMSAQPLNSQVEDEVDDQSGILSERIRLARHILPLS 115  
 Db 327 EDMQFASPYYIMSAQPLNSQVEDEVDDQSGILSERIRLARHILPLS 381

RESULT 6

ID Q8SX14 PRELIMINARY; PRT; 666 AA.  
 AC Q8SX14;  
 DT 01-JUN-2002 (TREMBrel. 21, Created)  
 DT 01-JUN-2002 (TREMBrel. 21, Last sequence update)  
 DT 01-OCT-2003 (TREMBrel. 25, Last annotation update)  
 DR RE5033P.  
 DE CG0027.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephioidea; Drosophilidae; Drosophila.  
 OC NCBI\_TAXID:7227;

Qy 1 DTFADSLRVNKILNSKGFSRKPVPHMPIDRIVQELQDMPPZEFKTSFHKYRH 60  
 Db 348 DIVSHSLATINMLNRAVGFKARHVLALVGFLIDKDVTEAMORRFHQIQLDTAHQRTRAP 407  
 Qy 61 EDMQFASPYYIMSAQPLNSQVEDEVDDQSGILSERIRLARHILPLS 120  
 Db 408 TDQYQAYAYSFLMSKETKWSVEIFEBFDTGSATSDREVRTFLTRIOPPLDSAMR 467

Qy 121 GLEMMLINGSK---MUPADITOLNNRPTQSYDPNLPPTKSYDUTNCKEVTDKHK 176  
 Db 468 YFEVWVCTRNLGMHLKDVTWHSYL--VYERVEDSNLPTITRDLVRCPLIAALAN 525

Qy 177 YDKKMKRKEIMGE--EEIAFKSIRTVSHVQJDIDKRNPKFVLANDDH\_KDA 233  
 Db 526 FAVERPKNFHVSKPRTSISNFMLTSNTEVESLDLRRNRKFGENDNDLADNGEDN 585

Qy 234 QTVAQAVLRDFYEEMFPSQFELPREYRNPLHMELOQEWRAVYRDKUKFWTHCVLATLIM 293  
 Db 586 EMTRHLADFLYSPPRRSKFEIPLPQTNRFSWRDIFQRMR-KRAVLVIGYVSLLL 644

Qy 294 FTFSFSAFQBLALKRKPFPR 314  
 Db 645 VCLDRFMCHHKAKLVRCVQR 665

RESULT 7

ID Q9V553 PRELIMINARY; PRT; 666 AA.  
 AC Q9V553;  
 DT 01-MAY-2000 (TREMBrel. 13, Created)  
 DT 01-OCT-2003 (TREMBrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TREMBrel. 25, Last annotation update)  
 DR CG0027.  
 DE CG0027.  
 OS Drosophila melanogaster (Fruit fly).

OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;	29.2%; Score 506; DB 5; Length 666;
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	Best Local Similarity 33.6%; Pred. No. 1.2e-31; Mismatches 133; Indels 10; Gaps 5;
OC	Ephydriidea; Drosophilidae; Drosophila.	Matches 108; Conservative 70; Mismatches 133; Indels 10; Gaps 5;
NCBI_TAXID=7227;		
[1]		
RP	SEQUENCE FROM N.A.	
RX	Medline-20160605; Pubmed=10731132;	
RA	Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,	
RA	Balilew R.M., Basl A., Baxendale J., Bayraktaroglu L., Beasley E.M.,	
RA	Benson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolhakov S.,	
RA	Borkova D., Botchar M.R., Bouck J., Brokstein P., Bottie P.,	
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,	
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,	
RA	Dobos J.F., Doyle C., Baxter E.G., Heit G., Nelson C.R., Gabor G.L.,	
RA	Dodson K., Douc L.E., Dowles M., Duana-Rocha S., Dunkov B.C., Dunn P.,	
RA	Durbin R.J., Evangelista C.C., Ferriz C., Ferreira S., Fleischmann W.,	
RA	Foster C., Gabril A.E., Garg N.M., Gelbart W.M., Glaser K.,	
RA	Gloedek A., Gong F., Gorrell J.H., Gu Z., Guan F., Harris M.,	
RA	Harris N.L., Harvey D., Heyman T.J., Hernandez J.R., Houck J.,	
RA	Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,	
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kerchum K.A.,	
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,	
RA	Lasko P., Lei Y., Levitt A.A., Li J., Li Z., Liang Y., Lin X.,	
RA	Liu X., Matrei B., McIntosh T.C., McLeod M.P., McPherson D.,	
RA	Merkulov G., Milashina N.V., Mobarry C., Morris J., Mosirafai A.,	
RA	Mount S.M., Moy M., Murphy B., Murphy J., Muzny D.M., Nelson D.L.,	
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacieb J.M.,	
RA	Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,	
RA	Reinert K., Remington K., Saunders R.D., Scheeler F., Sheen R.,	
RA	Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,	
RA	Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,	
RA	Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,	
RA	Wang Z.Y., Wasserman D.A., Weinstock G.M., Weissbach J.,	
RA	Williams R.M., Woodager, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,	
RA	Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,	
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,	
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C., Wei Z.,	
RT	"The genome sequence of <i>Drosophila melanogaster</i> .";	
RT	Science 287:2185-2195(2000).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RA	Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,	
RA	Hradecky Y., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,	
RA	Tupy J.L., Bergman C.M., Berman B.P., Carlson J.W., Celinker S.E.,	
RA	Clamp M.E., Drysdale R.A., Emmert D., Fife E., de Grey A.D.N.J.,	
RA	Harris N.L., Drymoniller B., Marshall B., Millburn G.H., Richter J.,	
RA	Russo S., Searle S.M.J., Smith E., Shiu S., Smutnicki F.,	
RA	Whitelock B.J., Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J.,	
RA	Lewis S.E.,	
RT	"Annotation of <i>Drosophila melanogaster</i> genome.";	
RT	Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.	
RN	[3]	
RP	SEQUENCE FROM N.A.	
RA	Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.	
RA	RAFB003834; AF585867.2; -.	
DR	GO: GO-0016020; C-membrane; IEA.	
DR	GO: GO-0030154; P:cell differentiation; IEA.	
DR	InterPro: IPR000000; Notch_dom.	
DR	Prmt; PR0006; notch; 1.	
DR	SMART; SM00004; NL; 1.	
SEQUENCE	666 AA; 77731 MW; EA232EC57C754FFF6 CRC64;	
QY		
Db	1 DTFAFDLSRVKNUKNSKGFTSRKPVAPMHPMTDRIVQELDMPEPEFDKTSFHVRHS 60	QY
Db	348 DIVSHSLATINMLNRANGKARHVLAHGFLJDIVEMQRBRHQDQLTAPAHQRFRAP 407	QY
Db	61 EDMQFASFSFYIYMSAVQPLNTSQVFDPEVDTQSGVSQDRETRTLATRHEPLSFDLT 120	QY
Db	408 TDLOQAFAYTSFLMSNETKMSWEIFDFDTQSATWSDREVTFETRIQPQLDMSMR 467	QY
Db	121 GLENHLINGSK---MLRADITOLNNTPTOSSYYDNLPPTKSLVYNCKEVTDKHK 176	QY
Db	468 YFEEVQONCTRNLQMLKUDTVHSTL--VYBREDSNLPPTTRDLYVRCPILAEALAN 525	QY
Db	177 YKDKNKXREIMGB--EELAFKQIRTVSHVUGQDQDQKRNPKFCVINDNDHEN-KDA 233	QY
Db	526 FAVPKYNFHVSPKRTSHSNFWMILTSNLTEVVESLDRLURRNPKFCINDNDANGEDN 585	QY
Db	234 QTIVGAVIRDYFYESMPPTSQPELPREYKRFPLMHEQEWRAYRDULKFWTHCVLATIM 293	QY
Db	586 EMVRHILBDPFYLSPPPRPSKFLPPQYRNRFESWRDFORWKR-RKAVLVIGYGVILLV 644	QY
Db	294 FTIFSSPPAEQLIAALKKFFP 314	QY
Db	645 VCLLRFMCHHKAKLVRRCVQ 665	QY
RESULT	8	
ID	068215	PRELIMINARY;
ID	068215	PRT;
AC	068215;	545 AA.
AC	068215;	
DT	01-AUG-1998 (TREMBLrel. 07, Created)	
DT	01-AUG-1998 (TREMBLrel. 07, Last sequence update)	
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)	
DR	SAB:	
DR	SACB:	
DR	Neisseria meningitidis (serogroup A)	
OS	Neisseria meningitidis (serogroup A); Betaproteobacteria; Neisseriales; OC	
OC	Neisseriaceae; Neisseria.	
OX	NCBI_TAXID=56599;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=82229;	
RX	Medline-98175678; Pubmed=9515923;	
RA	Swartley J.S., Liu L.-J., Miller Y.K., Martin L.E., Edupuganti S.,	
RA	Stephens D.S.,	
RT	"Characterization of the gene cassette required for biosynthesis of the (alpha-1->6)-linked N-acetyl-D-mannosamine-1-phosphate capsule of serogroup A <i>Neisseria meningitidis</i> ";	
RT	serogroup A <i>Neisseria meningitidis</i> ; ";	
RT	serogroup A <i>Neisseria meningitidis</i> ; "	
RT	serogroup A <i>Neisseria meningitidis</i> ; "	
RT	serogroup A <i>Neisseria meningitidis</i> ; "	
RT	serogroup A <i>Neisseria meningitidis</i> ; "	
RT	serogroup A <i>Neisseria meningitidis</i> ; "	
RT	serogroup A <i>Neisseria meningitidis</i> ; "	
RT	serogroup A <i>Neisseria meningitidis</i> ; "	
DR	EMBL; AR019760; AAC32861; -	
DR	EMBL; AR019760; AAC32861; -	
SQ	SEQUENCE 545 AA; 64097 MW; 1BC3400C5CC33DCD CRC64;	
Query	6.6%; Score 113.5; DB 2; Length 545;	
Match	Best Local Similarity 18.4%; Pred. No. 1.1; Matches 47; Conservative 36; Mismatches 65; Indels 107; Gaps 8;	
Match	6.6%; Score 113.5; DB 2; Length 545;	
Match	Best Local Similarity 18.4%; Pred. No. 1.1; Matches 47; Conservative 36; Mismatches 65; Indels 107; Gaps 8;	
QY	3 FADPSLRYKNIKMSKGFTSRKPVAPMHPMTDRIVQELDMPEPEFDKTSFHVRHS 62	QY
Db	395 YLNGARNANTLLEKEFKKFTKLHTHSPOSMDTDLFEMEKYPERKFNRFLNKFRSLDD 454	QY
QY	63 MQF-APSFYIYMSAVQPLNTSQVFDPEVDTQSGVSQDRETRTLATRHEPLSFDLT 121	QY
Db	455 IAVTGVLXHYY-----ALSGRALQS-----SDKTE 480	QY
QY	122 LEHMLINKSMLPADTOLANNPP-TOSYVPPNLPVPTKSLVNVCKPVTDKIHKA YDK 180	QY
Db	481 LVQQNHDFKK-----KLNNVWTLCERNFD-KLP----- 508	QY
QY	181 NKYRFEEIMEETTAFFMIRTNVSHVVGQDDIRKNRKFCVCLNDNDHNDKDAQTYKAVL 240	QY

Db	509	-----	LSVCINDGADSHLNEENWVQVI-	530
OY	241	RDFYESMPPIPOFE	255	SQ EMBL; AV234204; AA085302.1; -. 016B34F50DD5DD71 CRC64;
AC	084C00,	PRELIMINARY;	PRT;	SEQUENCE 545 AA; 61128 MW; 016B34F50DD5DD71 CRC64;
DT	01-JUN-2003	(TREMBrel. 24, Created)		Query Match 6.6%; Score 113.5; DB 2; Length 545;
DT	01-JUN-2003	(TREMBrel. 24, Last sequence update)		Best Local Similarity 18.4%; Pred. No. 1.1;
DT	01-JUN-2003	(TREMBrel. 24, Last annotation update)		Mismatches 36; Indels 107; Gaps 8;
DB	SACB.			Matches 47; Conservative 36; Mismatches 65; Indels 107; Gaps 8;
GN	Neisseria meningitidis			Query 3 FANSIRYNNKILNSKGFTSRKVPAHMPHMIDRIVKQELQDMFPEEDDKTFHKVRHSED 62
OS	SACB.			Db 395 YLNGARNANTLLKEFKKFTKHTHQPSMRTDILKEMEKYPPEENRTHLKNFRSLDD 454
OC	Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales; Neisseriaceae; Neisseria.			QY 63 MOF-AFSYYIYMSAVQPLNTISQVFDEVTDQSGVLSRETRTLATRHELPISLQLDTG 121
OC	NCBI_TaxID=487;			Db 455 IAVIGLYIHY-----ALLSGRALOS-----SDKTE 480
OX	[1]			QY 122 LEHMLINCSKMLPADITOLANNIPP-TQESYYDPNLPPVTSLVTNCKPVTDKIHAKYDK 180
RP	SBQUENCE FROM N.A.			Db 481 LVQONHDFKK-----KLANNVVLTKERNFD-KLP----- 508
RC	STRAIN=M4775;			QY 181 NKRFEEFGEEEAFKMRINTNISHWVQQLDDIRKRNPKFVCLNDNDHHNKDAQTWKAU 240
RA	Sacchi C.T., Whitney A.A., Mayer L.W., Mothershed E., Popovic T.;			: :
RT	"Neisseria meningitidis sacB gene.";			Db 509 ----- LSVCINDGADSHLNEENWVQVI- 530
RL	Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.			QY 241 RDFYESMPPIPOFE 255
DR	EMBL; AV234202; AA085300.1; -			Db 531 -KELETIPLPSSFE 544
SQ	SEQUENCE 545 AA; 61134 MW; A3648A3C85AA946F CRC64;			
Query Match	6.6%	Score 113.5;	DB 2;	Length 545;
Best Local Similarity	18.4%	Pred. No.	1.1;	
Matches	47;	Conservative	36;	
		Mismatches	65;	
		Indels	107;	Gaps 8;
QY	3 FADSIRYNNKILNSKGFTSRKVPAHMPHMIDRIVKQELQDMFPEEDDKTFHKVRHSED 62			
AC	084C00,	PRELIMINARY;	PRT;	
DT	01-JUN-2003	(TREMBrel. 24, Created)		
DT	01-OCT-2003	(TREMBrel. 25, Last sequence update)		
DB	4395 YLNGARNANTLLKEFKKFTKHTHQPSMRTDILKEMEKYPPEENRTHLKNFRSLDD 454			
GN	481 LVQONHDFKK-----KLANNVVLTKERNFD-KLP----- 508			
QY	63 MOF-AFSYYIYMSAVQPLNTISQVFDEVTDQSGVLSRETRTLATRHELPISLQLDTG 121			
DB	181 NKRFEEFGEEEAFKMRINTNISHWVQQLDDIRKRNPKFVCLNDNDHHNKDAQTWKAU 240			
DR	: :			
OC	455 IAVIGLYIHY-----ALLSGRALOS-----SDKTE 480			
OC	122 LEHMLINCSKMLPADITOLANNIPP-TQESYYDPNLPPVTSLVTNCKPVTDKIHAKYDK 180			
QY	481 LVQONHDFKK-----KLANNVVLTKERNFD-KLP----- 508			
DR	SBQUENCE FROM N.A.			
QY	STRAIN=M4775;			
AC	Sacchi C.T., Whitney A.A., Mayer L.W., Mothershed E., Popovic T.;			
AC	"Neisseria meningitidis sacB gene.";			
AC	Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AV234203; AA085301.1; -			
DR	SEQUENCE FROM N.A.			
QY	241 RDFYESMPPIPOFE 255			
DR	531 -KELETIPLPSSFE 544			
RESULT 10				
ID	084CZ9	PRELIMINARY;	PRT;	545 AA.
AC	084CZ9,	PRELIMINARY;	PRT;	545 AA.
DT	01-JUN-2003	(TREMBrel. 24, Created)		Query Match 6.6%; Score 113.5; DB 2; Length 545;
DT	01-JUN-2003	(TREMBrel. 24, Last sequence update)		Best Local Similarity 18.4%; Pred. No. 1.1;
DB	SacB.			Mismatches 36; Indels 107; Gaps 8;
GN	SACB.			Matches 47; Conservative 36; Mismatches 65; Indels 107; Gaps 8;
OS	Neisseria meningitidis			Query 3 FADSIRYNNKILNSKGFTSRKVPAHMPHMIDRIVKQELQDMFPEEDDKTFHKVRHSED 62
OC	Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales; Neisseriaceae; Neisseria.			Db 395 YLNGARNANTLLKEFKKFTKHTHQPSMRTDILKEMEKYPPEENRTHLKNFRSLDD 454
OC	NCBI_TaxID=487;			QY 63 MOF-AFSYYIYMSAVQPLNTISQVFDEVTDQSGVLSRETRTLATRHELPISLQLDTG 121
RN	[1]			Db 455 IAVIGLYIHY-----ALLSGRALOS-----SDKTE 480
RP	SBQUENCE FROM N.A.			Db 122 LEHMLINCSKMLPADITOLANNIPP-TQESYYDPNLPPVTSLVTNCKPVTDKIHAKYDK 180
RC	STRAIN=M2677;			QY 481 LVQONHDFKK-----KLANNVVLTKERNFD-KLP----- 508
RC	Sacchi C.T., Whitney A.A., Mayer L.W., Mothershed E., Popovic T.;			"Neisseria meningitidis sacB gene.";
RT	Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.			Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.
RL	EMBL; AV234203; AA085301.1; -			
DR	SEQUENCE FROM N.A.			
QY	STRAIN=M7060;			
AC	Sacchi C.T., Whitney A.M., Mayer L.W., Popovic T.;			
AC	"Neisseria meningitidis sacB gene.";			
AC	Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AV234203; AA085301.1; -			
DR	SEQUENCE FROM N.A.			
QY	STRAIN=M7060;			
AC	Sacchi C.T., Whitney A.M., Mayer L.W., Popovic T.;			
AC	"Neisseria meningitidis sacB gene.";			
AC	Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AV234203; AA085301.1; -			
DR	SEQUENCE FROM N.A.			
QY	STRAIN=M7060;			
AC	Sacchi C.T., Whitney A.M., Mayer L.W., Popovic T.;			
AC	"Neisseria meningitidis sacB gene.";			
AC	Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AV234203; AA085301.1; -			
DR	SEQUENCE FROM N.A.			
QY	STRAIN=M7060;			
AC	Sacchi C.T., Whitney A.M., Mayer L.W., Mothershed E., Popovic T.;			
AC	"Neisseria meningitidis sacB gene.";			
AC	Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AV234203; AA085301.1; -			
DR	SEQUENCE FROM N.A.			
QY	STRAIN=M7060;			
AC	Sacchi C.T., Whitney A.M., Mayer L.W., Mothershed E., Popovic T.;			
AC	"Neisseria meningitidis sacB gene.";			
AC	Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AV234203; AA085301.1; -			
DR	SEQUENCE FROM N.A.			
QY	STRAIN=M7060;			
AC	Sacchi C.T., Whitney A.M., Mayer L.W., Mothershed E., Popovic T.;			
AC	"Neisseria meningitidis sacB gene.";			
AC	Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AV234203; AA085301.1; -			
DR	SEQUENCE FROM N.A.			
QY	STRAIN=M7060;			
AC	Sacchi C.T., Whitney A.M., Mayer L.W., Mothershed E., Popovic T.;			
AC	"Neisseria meningitidis sacB gene.";			
AC	Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AV234203; AA085301.1; -			
DR	SEQUENCE FROM N.A.			
QY	STRAIN=M7060;			
AC	Sacchi C.T., Whitney A.M., Mayer L.W., Mothershed E., Popovic T.;			
AC	"Neisseria meningitidis sacB gene.";			
AC	Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AV234203; AA085301.1; -			
DR	SEQUENCE FROM N.A.			
QY	STRAIN=M7060;			
AC	Sacchi C.T., Whitney A.M., Mayer L.W., Mothershed E., Popovic T.;			
AC	"Neisseria meningitidis sacB gene.";			
AC	Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AV234203; AA085301.1; -			
DR	SEQUENCE FROM N.A.			
QY	STRAIN=M7060;			
AC	Sacchi C.T., Whitney A.M., Mayer L.W., Mothershed E., Popovic T.;			
AC	"Neisseria meningitidis sacB gene.";			
AC	Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AV234203; AA085301.1; -			
DR	SEQUENCE FROM N.A.			
QY	STRAIN=M7060;			
AC	Sacchi C.T., Whitney A.M., Mayer L.W., Mothershed E., Popovic T.;			
AC	"Neisseria meningitidis sacB gene.";			
AC	Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AV234203; AA085301.1; -			
DR	SEQUENCE FROM N.A.			
QY	STRAIN=M7060;			
AC	Sacchi C.T., Whitney A.M., Mayer L.W., Mothershed E., Popovic T.;			
AC	"Neisseria meningitidis sacB gene.";			
AC	Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AV234203; AA085301.1; -			
DR	SEQUENCE FROM N.A.			
QY	STRAIN=M7060;			
AC	Sacchi C.T., Whitney A.M., Mayer L.W., Mothershed E., Popovic T.;			
AC	"Neisseria meningitidis sacB gene.";			
AC	Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AV234203; AA085301.1; -			
DR	SEQUENCE FROM N.A.			
QY	STRAIN=M7060;			
AC	Sacchi C.T., Whitney A.M., Mayer L.W., Mothershed E., Popovic T.;			
AC	"Neisseria meningitidis sacB gene.";			
AC	Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AV234203; AA085301.1; -			
DR	SEQUENCE FROM N.A.			
QY	STRAIN=M7060;			
AC	Sacchi C.T., Whitney A.M., Mayer L.W., Mothershed E., Popovic T.;			
AC	"Neisseria meningitidis sacB gene.";			
AC	Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AV234203; AA085301.1; -			
DR	SEQUENCE FROM N.A.			
QY	STRAIN=M7060;			
AC	Sacchi C.T., Whitney A.M., Mayer L.W., Mothershed E., Popovic T.;			
AC	"Neisseria meningitidis sacB gene.";			
AC	Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AV234203; AA085301.1; -			
DR	SEQUENCE FROM N.A.			
QY	STRAIN=M7060;			
AC	Sacchi C.T., Whitney A.M., Mayer L.W., Mothershed E., Popovic T.;			
AC	"Neisseria meningitidis sacB gene.";			
AC	Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AV234203; AA085301.1; -			
DR	SEQUENCE FROM N.A.			
QY	STRAIN=M7060;			
AC	Sacchi C.T., Whitney A.M., Mayer L.W., Mothershed E., Popovic T.;			
AC	"Neisseria meningitidis sacB gene.";			
AC	Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AV234203; AA085301.1; -			
DR	SEQUENCE FROM N.A.			
QY	STRAIN=M7060;			
AC	Sacchi C.T., Whitney A.M., Mayer L.W., Mothershed E., Popovic T.;			
AC	"Neisseria meningitidis sacB gene.";			
AC	Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AV234203; AA085301.1; -			
DR	SEQUENCE FROM N.A.			
QY	STRAIN=M7060;			
AC	Sacchi C.T., Whitney A.M., Mayer L.W., Mothershed E., Popovic T.;			
AC	"Neisseria meningitidis sacB gene.";			
AC	Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AV234203; AA085301.1; -			
DR	SEQUENCE FROM N.A.			
QY	STRAIN=M7060;			
AC	Sacchi C.T., Whitney A.M., Mayer L.W., Mothershed E., Popovic T.;			
AC	"Neisseria meningitidis sacB gene.";			
AC	Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AV234203; AA085301.1; -			
DR	SEQUENCE FROM N.A.			
QY	STRAIN=M7060;			
AC	Sacchi C.T., Whitney A.M., Mayer L.W., Mothershed E., Popovic T.;			
AC	"Neisseria meningitidis sacB gene.";			
AC	Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AV234203; AA085301.1; -			
DR	SEQUENCE FROM N.A.			
QY	STRAIN=M7060;			
AC	Sacchi C.T., Whitney A.M., Mayer L.W., Mothershed E., Popovic T.;			
AC	"Neisseria meningitidis sacB gene.";			
AC	Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AV234203; AA085301.1; -			
DR	SEQUENCE FROM N.A.			
QY	STRAIN=M7060;			
AC	Sacchi C.T., Whitney A.M., Mayer L.W., Mothershed E., Popovic T.;			
AC	"Neisseria meningitidis sacB gene.";			
AC	Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AV234203; AA085301.1; -			
DR	SEQUENCE FROM N.A.			
QY	STRAIN=M7060;			
AC	Sacchi C.T., Whitney A.M., Mayer L.W., Mothershed E., Popovic T.;			
AC	"Neisseria meningitidis sacB gene.";			
AC	Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AV234203; AA085301.1; -			
DR	SEQUENCE FROM N.A.			
QY	STRAIN=M7060;			
AC	Sacchi C.T., Whitney A.M., Mayer L.W., Mothershed E., Popovic T.;			
AC	"Neisseria meningitidis sacB gene.";			
AC	Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AV234203; AA085301.1; -			
DR	SEQUENCE FROM N.A.			
QY	STRAIN=M7060;			
AC	Sacchi C.T., Whitney A.M., Mayer L.W., Mothershed E., Popovic T.;			
AC	"Neisseria meningitidis sacB gene.";			
AC	Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AV234203; AA085301.1; -			
DR	SEQUENCE FROM N.A.			
QY	STRAIN=M7060;			
AC	Sacchi C.T., Whitney A.M., Mayer L.W., Mothershed E., Popovic T.;			
AC	"Neisseria meningitidis sacB gene.";			
AC	Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AV234203; AA085301.1; -			
DR	SEQUENCE FROM N.A.			
QY	STRAIN=M7060;			
AC	Sacchi C.T., Whitney A.M., Mayer L.W., Mothershed E., Popovic T.;			
AC	"Neisseria meningitidis sacB gene.";			
AC	Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AV234203; AA085301.1; -			
DR	SEQUENCE FROM N.A.			
QY	STRAIN=M7060;			
AC	Sacchi C.T., Whitney A.M., Mayer L.W., Mothershed E., Popovic T.;			
AC	"Neisseria meningitidis sacB gene.";			
AC	Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AV234203; AA085301.1; -	</td		

QY	181 NKURFEIIMGEEETAPKMTIRTNVSHVVGOLDDIRKNPRKFVCLANDNDHNAQTYKAVL 240	DE Hypothetical protein (Fragment)
Db	509 -----LSVCINDGADSHNEEWVQVI- 530	OS Plasmodium falciparum (isolate 3D7).
AC	Q9WKB	OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
DT	01-OCT-2000 (TREMBrel. 15; Created)	NCBI_TaxID:56329;
DE	01-OCT-2000 (TREMBrel. 15; Last sequence update)	RN [1]
GN	SACB OR NMAB200.	RP SEQUENCE FROM N.A.
OS	Neisseria meningitidis (serogroup A).	RA STRAIN=D7;
OC	Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;	RA Alano P.; Elliott J.F.;
OC	Neisseriaceae; Neisseria.	RL Submitted (ARR-1993) to the EMBL/GenBank/DBJ databases.
OX	NCBI_TaxID=65699;	DR EMBL; MG9147; AA74653.1; -.
RP	SEQUENCE FROM N.A.	KW HYPOTHETICAL PROTEIN.
RC	STRAIN=22491 / Serogroup A / Serotype 4A;	FT HYPOTHETICAL PROTEIN.
RX	MEDLINE-20222556; PubMed=076119;	FT NON-TER 1
RA	Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,	FT 1048 AA; 126518 MW; 7AF051480FA22424 CRC64;
RA	Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,	SQ
RA	Davies R.M., Davis P., Devlin J., Feltwell T., Hamlin N., Holroyd S.,	Query Match 6.4%; Score 111; DB 5; Length 1048;
RA	Jigels K., Lehar S., Moules S., Mungall K., Quail M.A.,	Best Local Similarity 22.7%; Pred. No. 3 9;
RA	Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,	Matches 68; Conservative 50; Mismatches 94; Indels 88; Gaps 16;
RA	Whitehead S., Spratt B.G., Barrell B.J.,	Db 41 LQDMFPEEFDKSNSPHVKRSEDMOPAFSYFYILMSAVOPLNISQVDEVTDQSGVLSD- 99
RT	"Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491.";	Db 592 LEDIME--KKKYKEPKYSKERRISNLNLYEVNK---QKEHFKVDT---ISEH 639
RR	Nature 404:502-506 (2000).	Db 100 --RTR----TLATRHELPUSQ----DTGLEMILNSKSKMPADITOLNNPP-- 145
DR	EMBL; AU162752; CAB8514.1; -.	Db 640 KFOERQHMRDKLIENTHELYKENVYQIQIDLNYWHOLENHSIELQALOONKNIPRHL 699
DR	PIR; CAB2014; C82104.	Db 146 -----TOESYDPNLPV-TKSLVNCREPTIKHKAYD---KQYRPEI--- 187
KW	Complete proteome.	Db 700 NVLRKLEITTKRK---KOKPDKISTSHATDGGVSPTLIRGAHHNDITIGBDNDHVTLLI 757
SEQUENCE	545 AA; 64133 MW; 7D2564286C95CD3D CRC64;	Db 188 -----MGEBIAFKMIRTNVSHVQ-----LDIIRSPRKVCLNNDH 228
QY	Query Match 6.6%; Score 113.5; DB 16; Length 545;	Db 758 EQIOSKLTKGDNQNVGSTELEKNNISDQYLQLQDKLNVVEDIVKNLRNF---KHYIEK 814
Best Local Similarity 18.4%; Pred. No. 1, 1;	Matches 47; Conservative 36; Mismatches 65; Indels 107; Gaps 8;	Db 229 NHKDDQ-----TVKAYLDPFYESMIPPIPSQFELPREYRURFLHMELOEWRAYTDLK 281
AC	Q25857 PRELIMINARY; PRT; 3119 AA.	Db 815 LHKESSKINREKFTIKDVULSNVYSTL-----EYMKEL-LHDQEWSEFDELE 862
ID	Q25857	Db Q25857
AC	Q25857	DE Hypothetical protein (Fragment)
DT	01-NOV-1996 (TREMBrel. 01; Created)	OS Plasmodium falciparum (isolate 3D7).
DE	01-NOV-1996 (TREMBrel. 01; Last sequence update)	OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
GN	NCBI_TaxID=56329;	OX NCBI_TaxID=3D7;
DB	395 YLGARNVANTLSEKEFKETKLUHTHSFQSMDTDLFEMEKKYPEEPRNTLANKFRSLDD 454	RP SEQUENCE FROM N.A.
QY	63 MOP-AFSFTYVYLMVASQVPLNISQVFEDVTDQSGVSDREERTLATRINHPLSODLTG 121	RC STAIN=3D7;
DB	455 IAVTGILKHY-----ALISGRALQS-----SDRTE 480	RX MEDLINE-9630472; PubMed=8719156;
QY	122 LEHMLINGSKMILADITOLNNTPP-TOBSSYPPNLPVPTVKSLVNCVKPVTDKHKAYKD 180	RA Alano P., Read D., Bruce M., Akawa M., Kaido T., Tegoshi T., Elliott J.F.;
DB	481 LVQONHDFRK-----KLNNUVLTKERNFD-KLP----- 508	RA Bhatti S., Smith D.K., Luo C., Hansra S., Carter R., Elliott J.F.;
QY	181 NKRFREIIGEEETAPKMTIRTNVSHVVGOLDDIRKNPRKFVLDNNDHNAQTYKAVL 240	RA RT "COS cell expression cloning of Pf877, a Plasmodium falciparum gameteocyte antigen associated with Osmiophilic bodies.",
DB	509 -----LSVCINDGADSHNEEWVQVI- 530	RL Mol. Biochem. Parasitol. 74:143-156(1995).
QY	241 RDYFESMIPPIPSOBE 255	DR EMBL; AA37257.1; -.
DB	531 -KLEETLPLPSFE 544	DR PIR; T8414; T8414.
DR	InterPro; IPR00064; Crystallin.	DR PROSITE; PS00225; CRYSTALLIN-BTAGAMMA; 1.
SQ	SEQUENCE 3119 AA; 377353 MW; 51B99D25BDFCPC CRC64;	DR
RESULT	13	Query Match 6.4%; Score 111; DB 5; Length 3119;
Q26023	PRELIMINARY; PRT; 1048 AA.	Best Local Similarity 22.7%; Pred. No. 15; Mismatches 94; Indels 88; Gaps 16;
ID	Q26023	Matches 68; Conservative 50; Mismatches 94; Indels 88; Gaps 16;
AC	Q26023;	Db 41 LQMFPERFDKTSFKRHSDEMOPASFYVILMSAVQPLNISQVDEVTDQSGVLSD- 99
DT	01-NOV-1996 (TREMBrel. 01; Created)	Db 1261 LEIDIYME--KKKYKEPKYSKERRISNLNLYEVNK---QKEHFKVDT---ISEH 1308
DT	01-NOV-1996 (TREMBrel. 01; Last sequence update)	DT 01-OCT-2003 (TREMBrel. 25; Last annotation update)

QY	100	--REIR--TIA RILPLSLO-----DLG LEHMLINCSKQPADITQLNIPP-- 145	
Db	1309	KFQBIRQRMDKTIENIHELYKENTVQIQDILNTWYHOLENTISSELQALQONKNIPRL	1368
QY	146	-----TOESYDPNLPPV-TKSLVNCKPTDKHKAYD-----KKYRFI-- 187	
Db	1369	NVLEKLEITTKR-KKQPKDTSASHATDEQVSDTLRGAIENGHDGDIKGENDEVLLI	1426
QY	188	-----MGEERIAFKMVRTNVHGVQ-----LDDIKRKNPKFCVLNDNDIH	228
Db	1427	BQJOSLSKIKMGDNQVGSILEKUNLNSDQYQDQKLAVWEIYKNURNF-- KHYIE	1483
QY	229	NHKDQA-----TVKAVALDFYESMFPIPSQELPREVRNRFHLHMLQEWRAVDKUK	281
Db	1484	LHKESKINREKFITKVVDLUSVNST-----EYMKFL-LHDPOEWSFKEOBLE	1531
RESULT 15			
OI14T3	RP	PRELIMINARY; PRT; 3119 AA.	
ID	Q814T3	PRT;	
AC	Q114T3;		
DT	01-MAR-2003 (Tremblel, 23, Last sequence update)		
DT	01-OCT-2003 (Tremblel, 25, Last annotation update)		
DE	PRC377_protein.		
GN	PRC2405C.		
OS	Plasmodium falciparum (isolate 3D7).		
OX	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
RN	[1] NCBI_TaxId=36329;		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=3D7;		
RX	MEDLINE=22255705; PubMed=12368864;		
RA	Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W., Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K., Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S., Angiolillo S., Chan M.-S., Netea M., Shallal S.J., Suh B., Peterson J., Haff D., Mather M.W., Vaidya A.B., Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B., Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A., RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.R., Venter J.C., Carnici D.J., Hoffman S.L., Newbold C., Davis R.W., RA Fraser C.M., Barrell B., Hoffmann S.L., Newbold C., Davis R.W., RT Genome sequence of the human malaria parasite Plasmodium falciparum.", Nature 419:498-511(2002).		
RL	"Genome sequence of the human malaria parasite Plasmodium falciparum.", Nature 419:498-511(2002).		
DR	EMBL; AB014952; AAC36365.1; -		
DR	InterPro; IPR001664; Crystallin.		
DR	PROSITE; PS00225; CRYSTALLIN_BERGAMA; 1.		
SQ	SEQUENCE: PS00225; CRYSTALLIN_BERGAMA; 1. 3119 AA; 377380 MW; 3A6B67F1CD5CC1FF CRC64;		
Query Match 6.4%; Score 111; DB 5; Length 3119;			
Best Local Similarity 22.7%; Pred. No. 15; Matches 68; Conservative 50; Mismatches 94; Indels 88; Gaps 16;			
QY	41 LQDNFPEEPFDKTSFHKVRSEDMQFAFSYYFLMSAVQPLNTSQVDFEDVTQSGVLSD-	99	
Db	1261 LEDIYME--KKKYKEEYSKMRRISSNLNDYFVK-----QIKEHYKVDT-----ISEH	1308	
QY	100 --REIR-----TIA RILPLSLO-----DLG LEHMLINCSKQPADITQLNIP-- 145		
Db	1309 KFQBIRQRMDKTIENIHELYKENTVQIQDILNTWYHOLENTISSELQALQONKNIPRL	1368	
QY	146 -----TQESYDPNLPPV-TKSLVNCKPTDKHKAYD-----KKYRFI-- 187		
Db	1359 NVLEKLEITTKR-KKQPKDTSASHATDEQVSDTLRGAIENGHDGDIKGENDEVLLI	1426	
QY	188 -----MGEERIAFKMVRTNVHGVQ-----LDDIKRKNPKFCVLNDNDIH	228	
Db	1427 EQJOSLSKIKMGDNQVGSILEKUNLNSDQYQDQKLAVWEIYKNURNF-- KHYIE	1483	
QY	229 NHKDAO-----TVKAVALDFYESMFPIPSQELPREVRNRFHLHMLQEWRAVDKUK	281	

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2004, 11:01:28 ; Search time 28.1358 Seconds  
(without alignments)  
3062.893 Million cell updates/sec

Title: US-10-657-280-3

Perfect score: 1634

Sequence: 1 MAAGIARUILLIGISAGGPA.....TPRAKSPEQLRGDPGIRGSL 305

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 10%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04;\*

1: geneseqP1980s;\*

2: geneseqP1990s;\*

3: geneseqP2000s;\*

4: geneseqP2001s;\*

5: geneseqP2002s;\*

6: geneseqP2003as;\*

7: geneseqP2003bs;\*

8: geneseqP2004s;\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

### SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	1634	100.0	305	4 ABR61378
2	1634	100.0	305	5 ABGG92075
3	1634	100.0	305	6 ADA57553
4	1634	100.0	305	6 ADAS57217
5	1634	100.0	305	6 ADRA41096
6	1634	100.0	305	6 ADRA41445
7	1634	100.0	305	7 ABR91814
8	1634	100.0	305	7 ADR91665
9	1634	100.0	305	7 ADC74325
10	1634	100.0	305	7 ADC74555
11	1634	100.0	305	7 ADD27815
12	1634	100.0	305	7 ABW01490
13	1634	100.0	305	7 ADD37973
14	1634	100.0	305	7 ADD38034
15	1634	100.0	305	7 ADD80539
16	1634	100.0	305	7 ADD45631
17	1634	100.0	306	7 ADD45632
18	1634	100.0	306	6 ADL11624
19	1634	100.0	310	4 AAM25804
20	1288.5	78.9	307	4 ASR61381
21	1288.5	78.9	307	7 ADD27820
22	1288.5	78.9	307	7 ABW01493
23	1288.5	78.9	307	7 ABW01542
24	1244.5	76.2	296	2 AAY48273
25	1244.5	76.2	296	2 AAO30426
26	109.2	66.8	230	7 AAO30426

The invention relates to a novel isolated human N-acetylglucosamine-1-phototransferase (GlcNAc-phosphotranferase) (I) and phosphodiester

26	495.5	30.3	94	2 AAY73905	Human pro
27	171	10.5	647	3 AAG37770	Arabidops
28	171	10.5	647	3 AAG37715	Arabidops
29	168	10.3	350	3 AAG27509	Arabidops
30	168	10.3	358	3 AAG27509	Arabidops
31	155.5	9.5	597	7 ADB89035	TAX276.1
32	155.5	9.5	612	6 ABG74699	Human CGD
33	155.5	9.5	638	6 ABG74700	Human CGD
34	155.5	9.5	652	6 ABG74695	Human CGD
35	148	9.1	548	4 ABG62383	Drosophili
36	148	8.6	136	6 ABU70630	Human adi
37	140.5	8.6	524	4 AAU28043	Novel hum
38	140.5	8.6	527	4 AAU28042	Novel hum
39	140.5	8.6	527	4 AAB50266	Protein k
40	140.5	8.6	534	4 AAU28044	Novel hum
41	137.7	8.4	109	5 AAB66099	Wheat alp
42	125	7.6	483	4 AAB93456	Human pro
43	124.5	7.6	559	7 ADO8550	Novel pro
44	120.5	7.4	483	3 AAB08639	A murine
45	120.5	7.4	483	3 AAB08647	A murine

### ALIGNMENTS

RESULT 1					
ID	ABR61378	standard; protein; 305 AA.	XX	AC	ABR61378;
XX			XX		
DT			01-AUG-2003	(first entry)	
XX					Human GLCNAC-phosphotransferase gamma-subunit.
XX					Human; N-acetylglucosamine-1-phosphotransferase; nephrotropic;
XX					GlcNAc-phosphotranferase; phosphodiester alpha-N-Acetylglucosaminidase;
XX					N-acetylglucosamine-1-phosphodiester; phosphorylated lysosomal hydrolase;
OS					lysosomal storage disease; enzyme; gamma-subunit.
XX					Homo sapiens.
Key					location/qualifiers
Peptide					1. .24
FT					/label= Signal_peptide
FT					/label=.305
FT					/label=Mature_protein
Protein					

Novel N-acetylglucosamine-1-phosphotransferase and N-acetylglucosamine-1-phosphotriester alpha-N-Acetylglucosaminidase, useful for producing phosphorylated lysosomal hydrolase for treating lysosomal storage diseases.

Abw01539 GlcNAc-ph

Abw01493 Mouse pro

Abw01542 Mouse pro

Abw01490 Human Gic

Abw01491 Human sec

Abw01492 Human sec

Abw01493 Human Gic

Abw01494 Human sec

Abw01495 Human sec

Abw01496 Human sec

Abw01497 Human sec

Abw01498 Human sec

Abw01499 Human sec

Abw01500 Human sec

Abw01501 Human sec

Abw01502 Human sec

Abw01503 Human sec

Abw01504 Human sec

Abw01505 Human sec

Abw01506 Human sec

Abw01507 Human sec

Abw01508 Human sec

Abw01509 Human sec

Abw01510 Human sec

Abw01511 Human sec

Abw01512 Human sec

Abw01513 Human sec

Abw01514 Human sec

Abw01515 Human sec

Abw01516 Human sec

Abw01517 Human sec

Abw01518 Human sec

Abw01519 Human sec

Abw01520 Human sec

Abw01521 Human sec

Abw01522 Human sec

Abw01523 Human sec

Abw01524 Human nov

Abw01525 Human pro

Abw01526 Human pro

Abw01527 Human pro

Abw01528 Murine G1

Abw01529 GlcNAc-ph

Abw01530 GlcNAc-ph

Abw01531 Human Pro

Abw01532 Human Gic

Abw01533 Human sec

Abw01534 Human sec

Abw01535 Human sec

Abw01536 Human sec

Abw01537 Human sec

Abw01538 Human sec

Abw01539 Human sec

Abw01540 Human sec

Abw01541 Human sec

Abw01542 Human sec

Abw01543 Human sec

Abw01544 Human sec

Abw01545 Human sec

Abw01546 Human sec

Abw01547 Human sec

Abw01548 Human sec

Abw01549 Human sec

Abw01550 Human sec

Abw01551 Human sec

Abw01552 Human sec

Abw01553 Human sec

Abw01554 Human sec

Abw01555 Human sec

Abw01556 Human sec

Abw01557 Human sec

Abw01558 Human sec

Abw01559 Human sec

Abw01560 Human sec

Abw01561 Human sec

Abw01562 Human sec

Abw01563 Human sec

Abw01564 Human sec

Abw01565 Human sec

Abw01566 Human sec

Abw01567 Human sec

Abw01568 Human sec

Abw01569 Human sec

Abw01570 Human sec

Abw01571 Human sec

Abw01572 Human sec

Abw01573 Human sec

Abw01574 Human sec

Abw01575 Human sec

Abw01576 Human sec

Abw01577 Human sec

Abw01578 Human sec

Abw01579 Human sec

Abw01580 Human sec

Abw01581 Human sec

Abw01582 Human sec

Abw01583 Human sec

Abw01584 Human sec

Abw01585 Human sec

Abw01586 Human sec

Abw01587 Human sec

Abw01588 Human sec

Abw01589 Human sec

Abw01590 Human sec

Abw01591 Human sec

Abw01592 Human sec

Abw01593 Human sec

Abw01594 Human sec

Abw01595 Human sec

Abw01596 Human sec

Abw01597 Human sec

Abw01598 Human sec

Abw01599 Human sec

Abw01600 Human sec

Abw01601 Human sec

Abw01602 Human sec

Abw01603 Human sec

Abw01604 Human sec

Abw01605 Human sec

Abw01606 Human sec

Abw01607 Human sec

Abw01608 Human sec

Abw01609 Human sec

Abw01610 Human sec

Abw01611 Human sec

Abw01612 Human sec

Abw01613 Human sec

Abw01614 Human sec

Abw01615 Human sec

Abw01616 Human sec

Abw01617 Human sec

Abw01618 Human sec

Abw01619 Human sec

Abw01620 Human sec

Abw01621 Human sec

Abw01622 Human sec

Abw01623 Human sec

Abw01624 Human nov

Abw01625 Human pro

CC alpha-GlcNAcase (N-acetylglucosamine-1-phosphodiester alpha-N-  
 CC Acetylglucosaminidase) (III). The protein of the invention has  
 CC nephrotropic activity, and may be useful in enzyme replacement therapy. A  
 CC protein of the invention (I), (II) is useful for preparing a  
 CC phosphorylated lysosomal hydrolase. The phosphorylated hydrolase  
 CC comprising a terminal mannose-6-phosphate is useful for treating a  
 patient suffering from a lysosomal storage disease. The present sequence  
 CC is used in the exemplification of the invention.

XX Sequence 305 AA;

SQ

Query Match	100.0%	Score 1634;	DB 4;	Length 305;
Best Local Similarity	100.0%;	Pred. No. 3.e-161;	Mismatches 0;	Indels 0;
Matches 305;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 1 MAAGLARILLLGIGSAGGAPAGAAKOMVVEEPNAFGNNPFLPOASRLQAKRDPSPVSG 60

Db 1 MAAGLARILLLGIGSAGGAPAGAAKOMVVEEPNAFGNNPFLPOASRLQAKRDPSPVSG 60

QY 51 PVLHFRLSGKCFSLVESTYKYBPCPFRNTQHOTFRWNAFGSGILGIWHENBIANTFTG 120

Db 61 PVLHFRLSGKCFSLVESTYKYBPCPFRNTQHOTFRWNAFGSGILGIWHENBIANTFTG 120

QY 121 MWYRGDAGCRSRQRSKVLAACKSNRAHTSEPSCTWYALTFETPLVCHPHALLVPTL 180

Db 121 MWYRGDAGCRSRQRSKVLAACKSNRAHTSEPSCTWYALTFETPLVCHPHALLVPTL 180

QY 181 PEALQRQDQEVDLADLITQHOGHEKLRTLFEDAGLYKTPEENPTQLEGGPDLSIGFE 240

Db 181 PEALQRQDQEVDLADLITQHOGHEKLRTLFEDAGLYKTPEENPTQLEGGPDLSIGFE 240

QY 241 TLNCRKAHKELSKERIKLKGLTQHGPYTRPTESNLHGHETPRAKSPEQLRGDPG 300

Db 241 TLNCRKAHKELSKERIKLKGLTQHGPYTRPTESNLHGHETPRAKSPEQLRGDPG 300

QY 301 LRESL 305

Db 301 LRESL 305

RESULT 2

ID ABG92075 standard; protein: 305 AA.

ID ABG92075;

AC ABG92075;

DT 29-NOV-2002 (first entry)

XX Human receptors and membrane associated protein REMAP-18.

XX Human; receptor and membrane-associated protein; REMAP; atherosclerosis; cardiovascular disorder; hypertension; congestive heart failure; oedema; aneurysm; angina pectoris; ischaemic heart disease; lung disease; nausea; rheumatic heart disease; chronic obstructive pulmonary disease; anaemia; emphysema; bronchitis; gastrointestinal disorder; peptic ulcer; asthma; Crohn's disease; lipid metabolism disorder; Fabry's disease; dementia; pancreatitis; diabetes mellitus; hyperlipidaemia; autoimmune disorder; inflammatory disorder; neurological disorder; Cretzfeldt-Jakob disease; developmental disorder; protein replacement therapy; adenocarcinoma; developmental disorder; metabolic disorder; Alzheimer's disease; stroke; Parkinson's disease; multiple sclerosis; Cushing's syndrome; anxiety; schizophrenia; Addison's disease; endocrine disorder; gene therapy; gout; amnesia; cancer; leukemia; lymphoma; myeloma; sarcoma.

OS Homo sapiens.

PN WO200203006-A2.

XX 05-FEB-2002; 2002WO-US003868.

XX 06-FEB-2001; 2001US-0267201P.

PR 16-FEB-2001; 2001US-0269580P.

PR 09-APR-2001; 2001US-0282679P.

PR 02-MAY-2001; 2001US-0288295P.

PR 14-JAN-2002; 2002US-0348687P.

XX (INVY-) INCYTE GENOMICS INC.

PI Azinaii Y, Yue H, Ding L, Nguyen DB, Gandhi AR, Burford N; Tang TY; Thangavelu K, Elliott VS, Ramkumar J, Yao MG, Lal PG, Honchell CD; Swarnkar A, Warren BA, Walla NK, Policky JL, Xu Y; Au Young J, Baughn MR, Duggan BM, Lu DAM, Gietzen KU, Hillmann JL; PI Raumann BE, Lu Y, Sapperstein SK, Tran UK, Richardson TW; Emerling BM, Hafalia AWA, Burill JD, Marcus GA, Zingler KA; Kable AE, Gorvard AE;

XX DR WPT; 2002-62759167.

DR N-PDB; ABS67796.

XX PT New human receptor and membrane-associated proteins and polynucleotides, useful for diagnosing, treating or preventing cardiovascular disorders (e.g. aneurysms), neurological disorders (e.g. Parkinson's disease) or cancers.

PT PS Claim 2; Page 185-186; 262pp; English.

CC The present invention relates to a new receptor and membrane-associated protein (REMAP). The polypeptide, polynucleotide and agonist are useful for treating a condition associated with decreased expression of functional REMAP. The antagonist is useful for treating a disease associated with overexpression of functional REMAP. The anti-REMAP antibody is useful for diagnosing a condition or disease associated with the expression of REMAP. These polypeptides, polynucleotides, agonists and antagonists are particularly useful for diagnosing, treating or preventing cardiovascular (e.g. atherosclerosis, hypertension, aneurysms, congestive heart failure, arrhythmia, pectoralis, or ischaemic or rheumatic heart disease), lung (e.g. oedema, chronic obstructive pulmonary disease, emphysema or bronchitis), gastrointestinal (e.g. nausea, peptic ulcer or Crohn's disease), lipid metabolism (e.g. Fabry's disease, diabetes mellitus or hyperlipidaemia), autoimmune/inflammatory (e.g. anaemia, asthma, gout, pancreatitis or Crohn's disease), neurological (e.g. dementia, stroke, Alzheimer's disease, Creutzfeldt-Jakob disease, multiple sclerosis, Parkinson's disease, anxiety, schizophrenia or amnesia), metabolic (e.g. Addison's disease), developmental (e.g. Cushing's syndrome), endocrine or cell proliferative disorders (e.g. cancers including adenocarcinoma, leukaemia, lymphoma, myeloma or sarcoma). The present amino acid sequence represents a human REMAP protein of the invention.

CC SQ Sequence 305 AA;

Query Match	100.0%	Score 1634;	DB 5;	Length 305;
Best Local Similarity	100.0%;	Pred. No. 3.e-161;	Mismatches 0;	Indels 0;
Matches 305;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 1 MAAGLARILLLGIGSAGGAPAGAAKOMVVEEPNAFGNNPFLPOASRLQAKRDPSPVSG 60

Db 1 MAAGLARILLLGIGSAGGAPAGAAKOMVVEEPNAFGNNPFLPOASRLQAKRDPSPVSG 60

QY 61 PVLHFRLSGKCFSLVESTYKYBPCPFRNTQHOTFRWNAFGSGILGIWHENBIANTFTG 120

Db 61 PVLHFRLSGKCFSLVESTYKYBPCPFRNTQHOTFRWNAFGSGILGIWHENBIANTFTG 120

QY 121 MWYRGDAGCRSRQRSKVLAACKSNRAHTSEPSCTWYALTFETPLVCHPHALLVPTL 180

Db 121 MWYRGDAGCRSRQRSKVLAACKSNRAHTSEPSCTWYALTFETPLVCHPHALLVPTL 180

QY 181 PEALQRQDQEVDLADLITQHOGHEKLRTLFEDAGLYKTPEENPTQLEGGPDLSIGFE 240

Db 181 PEALQRQDQEVDLADLITQHOGHEKLRTLFEDAGLYKTPEENPTQLEGGPDLSIGFE 240

QY 241 TLNCRKAHKELSKERIKLKGLTQHGPYTRPTESNLHGHETPRAKSPEQLRGDPG 300

Db 241 TLNCRKAHKELSKERIKLKGLTQHGPYTRPTESNLHGHETPRAKSPEQLRGDPG 300

QY  
301 LRGSL 305  
|||||  
Db 301 LRGSL 305

RESULT 3  
ADA57553

CC gels, to raise antibodies, for testing biological activities, and for  
CC treating or preventing neural disorders, immune system disorders,  
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,  
CC renal, proliferative and/or cancerous diseases. This sequence corresponds  
CC to one of the polypeptide of the invention. Note: The sequence data for  
CC this patent did form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC fip.wipo.int/pdb/published\_pct\_sequences.

gels, to raise antibodies, for testing biological activities, and for treating or preventing neural disorders, immune system disorders, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal, proliferative and/or cancerous diseases. This sequence corresponds to one of the polypeptide of the invention. Note: The sequence data for this patent did form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp://wipo.int/pub/published\\_pct\\_sequences](ftp://wipo.int/pub/published_pct_sequences).

20-NOV-2003 (first entry)						
	Query	Match	Score	DB	Length	
Human secreted protein #500.			100.0%	1634	6	length 305;
	Best Local Similarity	100.0%	Score	1634	6	
	Matches	305;	DB	6	Length	305;
	Conservative	0;	Pred. No.	3.2e-161;		
	Mismatches	0;	No. of	0;		
	Indels	0;				
	Gaps	0;				
immunosuppressive; antiinflammatory; antiasthmatic; antiallergic; cytostatic; cerebroprotective; neuroprotective; nootropic; cardiovascular; antiatherosclerotic; gene therapy; human secreted protein; immune disorder; inflammation; respiratory disorder; cancer; CNS disorder; neurodegenerative disorders; inflammatory bowel disease; nephritis; Crohn's disease; asthma; allergy; multiple sclerosis; ischaemic brain injury; Parkinson's disease; Alzheimer's disease; atherosclerosis; Myocarditis; chromosome mapping; triple helix formation; antisense gene therapy; forensic biology.						
Homo sapiens.						

**PI** Rosen CA, Ruben SM;  
**XX**  
**DR** WPI; 2003-167512/16.  
**DR** N-PSDB; AD56321.  
**XX**  
**PT** New human secreted polypeptides and polynucleotides, useful for diagnosing, treating or preventing e.g. immune disorders, inflammatory conditions, respiratory disorders, cancers, CNS disorders, or neurodegenerative disorders.  
**XX**  
**RS** Claim 13; SEQ ID NO 1407; 1754pp; English.  
**XX**  
**CC** The invention relates to 592 new human secreted polypeptides useful for diagnosing, treating or preventing e.g. immune disorders, inflammatory conditions, respiratory disorders, cancers, CNS disorders, or neurodegenerative disorders, or polypeptides comprising a amino acid sequence at least 95% identical to the new sequences. The polypeptides, the polypeptides are useful in preparing diagnostic or pharmaceutical compositions for diagnosing, treating or preventing an e.g. immune disorders, inflammatory conditions (e.g. inflammatory bowel disease, nephritis or Crohn's disease), respiratory disorders (e.g. asthma and allergy), cancers (e.g. gastric, ovarian or lung cancer), CNS disorders (e.g. multiple sclerosis or ischaemic brain injury), neurodegenerative disorders (e.g. Parkinson's disease or Alzheimer's disease), and cardiovascular disorders (e.g. atherosclerosis or myocarditis). The polynucleotides are useful for chromosome identification, chromosome mapping, for controlling gene expression through triple helix formation or antisense DNA or RNA, in gene therapy, for identifying individuals from minute biological samples, in forensic biology, and as hybridization probes. The polypeptides are useful for as molecular weight markers on cellulose dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) gels, to raise antibodies, for testing biological activities, and for treating or preventing, neural disorders, immune system disorders, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal, proliferative and/or cancerous diseases. This sequence corresponds to one of the polypeptide of the invention. Note: The sequence data for this patent did form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp://wipo.int/pdb/published\\_pct\\_sequences](ftp://wipo.int/pdb/published_pct_sequences).  
**XX**  
**SQ** Sequence 305 AA;  
**Query Match** 100.0%; Score 1634; DB 6; Length 305;  
**Best Local Similarity** 100.0%; **Pred.** No. 3.2e-161; **Matches** 305; **Conservative** 0; **Mismatches** 0; **Indels** 0; **Gaps** 0;  
**CC** The invention relates to novel genes Ad439629-ADA40565 and proteins ADA40566-ADA41501 for human secreted proteins, useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The polypeptides, nucleic acid molecules, antibodies or their fragments, and agonists or antagonists that bind to the polypeptide are useful for preparing a diagnostic or pharmaceutical composition for diagnosing or treating cancer or other hyperproliferative disorder. The polypeptide and nucleic acid molecules are also useful for detecting, preventing, diagnosing, prognosticating, treating or ameliorating cancer or other hyperproliferative disorders including neoplasms, autoimmune disorders (e.g. diabetes, rheumatoid arthritis, systemic lupus erythematosus, multiple sclerosis, autoimmune thyroiditis or haemolytic anaemia), hematopoietic or haematological disorders (e.g. anaemia, thrombocytopenia), allergic reactions including asthma or eczema, inflammatory disorders (e.g. ischaemia-reperfusion injury, inflammatory bowel disease or Crohn's disease), neurodegenerative disorders (e.g. Alzheimer's disease or Parkinson's disease), cardiovascular disorders (e.g. atherosclerosis, myocarditis), infectious diseases (bacterial, fungal or viral infections including HIV/AIDS), or wound healing and disorders of epithelial cell proliferation. The nucleic acids are also useful for chromosome identification, radiation hybrid mapping or long-range restriction mapping, as molecular weight markers, or as hybridization or diagnostic probes. The polypeptides and antibodies are useful for providing immunological probes for differential identification of the tissues immunohistochemistry assays. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp://wipo.int/pdb/published\\_pct\\_sequences](ftp://wipo.int/pdb/published_pct_sequences).  
**XX**  
**RESULT** 5  
**ID** ADA41096  
**ID** ADA41096 standard; protein; 305 AA.  
**XX**  
**ID** ADA41096;  
**AC**  
**XX**  
**DT** 20-NOV-2003 (first entry)  
**XX**  
**DE** Human secreted protein.  
**XX**  
**KW** Human; secreted protein; cancer; hyperproliferative disorder; rheumatoid arthritis; autoimmune disorder; haemato poetic disorder; anaemia; allergic reaction; asthma; cardiovascular disorder; wound healing; cryostatic; immunosuppressive; nontropic; neuroprotective; antiviral; antiallergic; hepatotoxic; antidiabetic; antiinflammatory; vulnerability; cardiot; gene therapy.  
**XX**  
**KW** Homo sapiens.  
**XX**  
**OS** Homo sapiens.  
**XX**  
**PN** WO200212993-A2.  
**XX**  
**PD** 27-DEC-2002.  
**XX**  
**PP** 19-MAR-2002; 2002WO-US008123.  
**XX**  
**PR** 21-MAR-2001; 2001US-0277340P.  
**PR** 19-JUL-2001; 2001US-0306171P.  
**PR** 13-NOV-2001; 2001US-0331281P.  
**XX**  
**PA** (HOMA-) HUMAN GENOME SCI INC.  
**XX**  
**PA** Rosen CA, Ruben SM;  
**XX**  
**PI** DR  
**XX**  
**PT** WPI; 2003-175238/17.  
**XX**  
**PT** New human secreted proteins and nucleic acid molecules, useful for preparing a diagnostic or pharmaceutical composition for diagnosing, preventing or treating cancer or other hyperproliferative disorder, asthma, allergies or AIDS.  
**XX**  
**PS** Claim 1; SEQ ID NO 1478; 3205pp; English.  
**CC** The invention relates to novel genes Ad439629-ADA40565 and proteins ADA40566-ADA41501 for human secreted proteins, useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The polypeptides, nucleic acid molecules, antibodies or their fragments, and agonists or antagonists that bind to the polypeptide are useful for preparing a diagnostic or pharmaceutical composition for diagnosing or treating cancer or other hyperproliferative disorder. The polypeptide and nucleic acid molecules are also useful for detecting, preventing, diagnosing, prognosticating, treating or ameliorating cancer or other hyperproliferative disorders including neoplasms, autoimmune disorders (e.g. diabetes, rheumatoid arthritis, systemic lupus erythematosus, multiple sclerosis, autoimmune thyroiditis or haemolytic anaemia), hematopoietic or haematological disorders (e.g. anaemia, thrombocytopenia), allergic reactions including asthma or eczema, inflammatory disorders (e.g. ischaemia-reperfusion injury, inflammatory bowel disease or Crohn's disease), neurodegenerative disorders (e.g. Alzheimer's disease or Parkinson's disease), cardiovascular disorders (e.g. atherosclerosis, myocarditis), infectious diseases (bacterial, fungal or viral infections including HIV/AIDS), or wound healing and disorders of epithelial cell proliferation. The nucleic acids are also useful for chromosome identification, radiation hybrid mapping or long-range restriction mapping, as molecular weight markers, or as hybridization or diagnostic probes. The polypeptides and antibodies are useful for providing immunological probes for differential identification of the tissues immunohistochemistry assays. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp://wipo.int/pdb/published\\_pct\\_sequences](ftp://wipo.int/pdb/published_pct_sequences).  
**XX**

Query Match 100.0%; Score 1634; DB 6; Length 305;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-161; Mismatches 0; Index 0; Gaps 0;  
 Matches 305; Conservative 0; MisMatches 0; Indels 0; Gaps 0;

QY 1 MAQLARILLLGLSAGGPPAPAGAAKMVKEPNAPGUNPFLPQASRQAKRDPSVG 60  
 Db 1 MAQLARILLLGLSAGGPPAPAGAAKMVKEPNAPGUNPFLPQASRQAKRDPSVG 60

QY 61 PVLFLRISKGKPSLVLSTYKYEFCPFPNVTQHETQFRWNAYSGIGIWHWEIANNTFG 120  
 Db 61 PVLFLRISKGKPSLVLSTYKYEFCPFPNVTQHETQFRWNAYSGIGIWHWEIANNTFG 120

QY 121 MWRRDGACRSRSRQSRYKVELAGKSNLHAYSEPTCVCYALFEPVLCPHALVPTL 180  
 Db 121 MWRRDGACRSRSRQSRYKVELAGKSNLHAYSEPTCVCYALFEPVLCPHALVPTL 180

QY 121 MWRRDGACRSRSRQSRYKVELAGKSNLHAYSEPTCVCYALFEPVLCPHALVPTL 180  
 Db 121 MWRRDGACRSRSRQSRYKVELAGKSNLHAYSEPTCVCYALFEPVLCPHALVPTL 180

QY 181 PEALQRQMDQVEQDIALETLTPOGHKLRLTFLDAGYKTPPEENPTQEGPDSIGE 240  
 Db 181 PEALQRQMDQVEQDIALETLTPOGHKLRLTFLDAGYKTPPEENPTQEGPDSIGE 240

QY 241 TLNCRKAKESKEIRKLKGUTQHGPYKPRTESENLEHGHETPRAKSPEQLRGDPG 300  
 Db 241 TLNCRKAKESKEIRKLKGUTQHGPYKPRTESENLEHGHETPRAKSPEQLRGDPG 300

QY 301 LRSSL 305  
 Db 301 LRSSL 305

SQ

RESULT 6

ADA41445 standard; protein; 305 AA.

XX

AC ADA41445;

DT 20-NOV-2003 (first entry)

DE Human secreted protein.

XX

XX Human; secreted protein; cancer; hyperproliferative disorder; rheumatoid arthritis; autoimmune disorder; haemopoietic disorder; anaemia; allergic reaction; asthma; cardiovascular disorder; wound healing; cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral; antiarrhythmic; hepatotropic; antidiabetic; antiinflammatory; vulnerary; cardiant; gene therapy.

OS Homo sapiens.

XX WO2002102933-A2.

PN XX

PD 27-DEC-2002.

PR 21-MAR-2001; 2001US-0277340P.

PR 19-JUL-2001; 2001US-0306171P.

PR 13-NOV-2001; 2001US-0331287P.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PT Rosen CA., Ruben SM;

DR XX

WPI; 2003-175238/17.

XX

PT New human secreted proteins and nucleic acid molecules, useful for preparing a diagnostic or pharmaceutical composition for preventing or treating cancer or other hyperproliferative disorder, asthma, allergies or AIDS.

XX

CC Claim 1; SEQ ID NO 1828; 3205pb; English.

XX

CC The invention relates to novel genes ADA39629-ADA40565 and proteins

SQ Sequence 305 AA;

RESULT 7

A0B91814

ID A0B91814 standard; protein; 305 AA.

XX

AC A0B91814;

DT 04-DEC-2003 (first entry)

XX

DE Human secreted protein #SEQ ID 760.

XX

KW Secreted protein; gene therapy; antidiabetic; diabetes; human.

XX

OS Homo sapiens.

XX

ADA40566-ADA41501 for human secreted proteins, useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The polypeptides, nucleic acid molecules, antibodies or their fragments, and agonists or antagonists that bind to the polypeptide are useful for preparing a diagnostic or pharmaceutical composition for diagnosing or treating cancer or other hyperproliferative disorder. The polypeptides and nucleic acid molecules are also useful for detecting, preventing, diagnosing, prognosticating, treating or ameliorating cancer or other hyperproliferative disorders including neoplasms, autoimmune disorders (e.g. diabetes, rheumatoid arthritis, systemic lupus erythematosus, multiple sclerosis, autoimmune thyroiditis or haemolytic anaemia), haematopoietic or haematological disorders (e.g. anaemia, thrombocytopenia), allergic reactions including asthma or eczema, inflammatory disorders (e.g. ischaemia-reperfusion injury, inflammatory bowel disease or Crohn's disease), neurodegenerative disorders (e.g. Alzheimer's disease or Parkinson's disease), cardiovascular disorders (e.g. atherosclerosis, myocarditis), infectious diseases (bacterial, fungal or viral infections including HIV/AIDS), or wound healing, disorders of epithelial cell proliferation. The nucleic acids are also useful for chromosome identification, radiation hybrid mapping or long-range restriction mapping, as molecular weight markers or as hybridization or diagnostic probes. The polypeptides and antibodies are useful for providing immunological probes for differential identification of the tissues immunohistochemistry assays. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

PN WO2003004622-A2.  
 XX  
 PD 16-JAN-2003.  
 XX  
 PF 19-MAR-2002; 2002WO-US008124.  
 XX  
 PR 21-MAR-2001; 2001US-0377340P.  
 PR 13-JUL-2001; 2001US-0306171P.  
 PR 13-NOV-2001; 2001US-0331287P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PT Rosen CA, Ruben SM;  
 XX  
 DR WPI; 2003-2229407/22.  
 XX  
 PT Nucleic acid encoding a human secreted protein is useful in diagnosing or treating diabetes or conditions related to diabetes.  
 XX  
 PS Claim 3; SEQ ID NO 760; 1537pp; English.  
 XX  
 CC The invention relates to isolated nucleic acid molecules ADB91055-  
 CC ADB91448 and ADB91835-ADB91911 encoding human secreted proteins ADB91449-  
 CC ADB91834. Also disclosed is a recombinant vector comprising a polynucleotide of the invention, and a recombinant host cell comprising  
 CC the recombinant vector. The polypeptide of the invention is useful in identifying a binding partner by contacting the polypeptide with a binding partner, and determining whether the binding partner increases or decreases activity of the polypeptide. The polypeptide, polynucleotide, antibody or its fragment, agonist or antagonist are useful for preparing a pharmaceutical composition for diagnosing or treating diabetes or conditions related to diabetes. The present sequence is that of the human immunoglobulin Fc portion used to generate fusion proteins, increasing the stability of the fused protein as compared to the secreted protein only. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp://wipo.int/pub/published\\_pct\\_sequences](ftp://wipo.int/pub/published_pct_sequences).  
 XX  
 Sequence 305 AA:  
 SQ Query Match 100.0%; Score 1634; DB 7; Length 305;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-161; Mismatches 0; Indels 0; Gaps 0;  
 Matches 305; Conservative 0;  
 CC 1 MAAGLARILLLGSGAGCPAPAGAAKQKVVERPNAVNPNPLPQASRLQAKRDPSPVSG 60  
 1 MAAGLARILLLGSGAGCPAPAGAAKQKVVERPNAVNPNPLPQASRLQAKRDPSPVSG 60  
 DB  
 QY 61 PVHLFRLSGKCFSLVESTYKYFPCPFHNTQHOTFRWNAVSGIGIWHWEIANNTFTG 120  
 DB 61 PVHLFRLSGKCFSLVESTYKYFPCPFHNTQHOTFRWNAVSGIGIWHWEIANNTFTG 120  
 QY 121 MWYRGGDCRSRSRQSKVLEACKSNRLAHYSEPTCVALTFETPLVCHHALVYPTL 180  
 DB 121 MWYRGGDCRSRSRQSKVLEACKSNRLAHYSEPTCVALTFETPLVCHHALVYPTL 180  
 QY 181 PEALQRQMDQVEQDIALBLITPOGHKEKLRTLPEDAGYLKTPPEENPTQLEGGSPLIGFE 240  
 DB 181 PEALQRQMDQVEQDIALBLITPOGHKEKLRTLPEDAGYLKTPPEENPTQLEGGSPLIGFE 240  
 QY 241 TLENORKAHKELSKIEIKLKGLUTQHGYIPYRPTESNLIGHETPRAKSPEQLRGDPG 300  
 DB 241 TLENORKAHKELSKIEIKLKGLUTQHGYIPYRPTESNLIGHETPRAKSPEQLRGDPG 300  
 QY 301 LRGSI 305  
 DB 301 LRGSI 305  
 SQ Sequence 305 AA:  
 SQ Query Match 100.0%; Score 1634; DB 7; Length 305;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-161; Mismatches 0; Indels 0; Gaps 0;  
 Matches 305; Conservative 0;  
 CC 1 MAAGLARILLLGSGAGCPAPAGAAKQKVVERPNAVNPNPLPQASRLQAKRDPSPVSG 60  
 1 MAAGLARILLLGSGAGCPAPAGAAKQKVVERPNAVNPNPLPQASRLQAKRDPSPVSG 60  
 DB 61 PVHLFRLSGKCFSLVESTYKYFPCPFHNTQHOTFRWNAVSGIGIWHWEIANNTFTG 120  
 DB 61 PVHLFRLSGKCFSLVESTYKYFPCPFHNTQHOTFRWNAVSGIGIWHWEIANNTFTG 120  
 QY 121 MWYRGGDCRSRSRQSKVLEACKSNRLAHYSEPTCVALTFETPLVCHHALVYPTL 180  
 DB 121 MWYRGGDCRSRSRQSKVLEACKSNRLAHYSEPTCVALTFETPLVCHHALVYPTL 180  
 QY 181 PEALQRQMDQVEQDIALBLITPOGHKEKLRTLPEDAGYLKTPPEENPTQLEGGSPLIGFE 240  
 DB 181 PEALQRQMDQVEQDIALBLITPOGHKEKLRTLPEDAGYLKTPPEENPTQLEGGSPLIGFE 240  
 QY 241 TLENORKAHKELSKIEIKLKGLUTQHGYIPYRPTESNLIGHETPRAKSPEQLRGDPG 300  
 DB 241 TLENORKAHKELSKIEIKLKGLUTQHGYIPYRPTESNLIGHETPRAKSPEQLRGDPG 300  
 QY 241 TLENORKAHKELSKIEIKLKGLUTQHGYIPYRPTESNLIGHETPRAKSPEQLRGDPG 300  
 DB 241 TLENORKAHKELSKIEIKLKGLUTQHGYIPYRPTESNLIGHETPRAKSPEQLRGDPG 300

RESULT 8  
 ADB91665  
 ID ADB91665 standard; protein; 305 AA.  
 XX



PS

Claim 16; SEQ ID NO 1198; 2272pp; English.

XX  
CC The invention relates to a novel human secreted polypeptide comprising a defined sequence given in the specification. The polypeptide, nucleic acid molecule, antibody, agonist or antagonist of the invention may be used for preparing a composition for diagnosing or treating a

CC haemopoietic or hematologic disorder such as anaemia, autoimmune

CC disorders such as rheumatoid arthritis, inflammation, Grave's disease,

CC diabetes, systemic lupus erythematosus or glomerulonephritis, neurodegenerative disorders including Parkinson's disease and Alzheimer's

CC disease, wounds and hyperproliferative disorders including

CC atherosclerosis or cancer, as well as bacterial, viral, fungal or

CC parasitic infections. The polypeptide may also be used during gene

CC therapy procedures and for identifying a binding partner by contacting

CC the polypeptide with a binding partner and determining whether the

CC binding partner increases or decreases the activity of the polypeptide.

The current sequence is that of the human secreted protein of the

CC invention.

XX  
SQ Sequence 305 AA;

Query Match 100.0%; Score 1634; DB 7; Length 305;  
Best Local Similarity 100.0%; Pred. No. 3.2e-161; Mismatches 0; Indels 0; Gaps 0;

Matches 305; Conservative 0; MisMatches 0; Indels 0; Gaps 0;

DY 1 MAAGLARUILLGLISAGGPPAGAAKOKVVEPNAFGVNNPLPQSLQAKRDPSPVSG 60  
1 MAAGLARUILLGLISAGGPPAGAAKOKVVEPNAFGVNNPLPQSLQAKRDPSPVSG 60  
Dy 61 PVHLFRSGKCFSLVESTYKYERCPFVNTOHQETFRMNASGILGIWHENELANTFTG 120  
Db 61 PVHLFRSGKCFSLVESTYKYERCPFVNTOHQETFRMNASGILGIWHENELANTFTG 120  
QY 121 MWRRDGACRSRSRQRSKYELACKSNLAHSEPTCVALTFETLVCHHALVYPTL 180  
121 MWRRDGACRSRSRQRSKYELACKSNLAHSEPTCVALTFETLVCHHALVYPTL 180  
Dy 181 PEALQRQMDQEVQDADLITPOGHKLURTFEDAGYLIKTPBENPTQLEGGSIGFE 240  
181 PEALQRQMDQEVQDADLITPOGHKLURTFEDAGYLIKTPBENPTQLEGGSIGFE 240  
QY 241 TLENCRKAHKELSKEIKRLKGILTQHGIPYTRPTESNLLEHIGHETPRAKSPEQLRGDPG 300  
241 TLENCRKAHKELSKEIKRLKGILTQHGIPYTRPTESNLLEHIGHETPRAKSPEQLRGDPG 300  
Dy 301 LR GSL 305  
Db 301 LR GSL 305  
Dy 301 LR GSL 305  
Db 301 LR GSL 305

RESULT 11

ADD27815  
ID ADD27815 standard; protein; 305 AA.AC ADD27815;  
DT 15-JAN-2004 (first entry)

DE Human GlcNAc-phosphotransferase gamma subunit.  
XX  
KW human; protein phosphorylation; soluble GlcNAc-phosphotransferase;  
KW UDP-N-acetylglucosamine; lysosomal storage disease; Fabry's disease.  
XX  
OS Homo sapiens.  
PN US2003119088-A1.  
PD 26-JUN-2003.  
XX  
PF 21-DEC-2001; 2001US-00023888.  
XX  
PR 21-DEC-2001; 2001US-00023888.  
XX

PA (NOVA-) NOVAZYME PHARM INC.

PI Canfield W, Kudo M;

XX DR WPI: 2003-801323/75.

DR N-PSSB; ADD27814.

XX PT Phosphorylating a protein for treating a patient suffering from a

CC lysosomal storage disease e.g. Fabry's disease by contacting the protein

CC with a soluble GlcNAc-phosphotransferase and producing a phosphorylated

PT protein.

XX PS Claim 11; SEQ ID NO 7; 55pp; English.

CC The invention relates to a method of phosphorylating a protein comprising

CC contacting the protein with a soluble GlcNAc-phosphotransferase (UDP-N-

acetylglucosamine) and producing a phosphorylated protein. The method is

useful for treating a patient suffering from a lysosomal storage disease

e.g. Fabry's disease. The present sequence represents the amino acid

CC sequence of the human GlcNAc-phosphotransferase gamma subunit.

XX  
SQ Sequence 305 AA;

Query Match 100.0%; Score 1634; DB 7; Length 305;  
Best Local Similarity 100.0%; Pred. No. 3.2e-161; Mismatches 0; Indels 0; Gaps 0;

Matches 305; Conservative 0; MisMatches 0; Indels 0; Gaps 0;

Dy 1 MAAGLARUILLGLISAGGPPAGAAKOKVVEPNAFGVNNPLPQSLQAKRDPSPVSG 60  
1 MAAGLARUILLGLISAGGPPAGAAKOKVVEPNAFGVNNPLPQSLQAKRDPSPVSG 60  
Dy 61 PVHLFRSGKCFSLVESTYKYERCPFVNTOHQETFRMNASGILGIWHENELANTFTG 120  
Db 61 PVHLFRSGKCFSLVESTYKYERCPFVNTOHQETFRMNASGILGIWHENELANTFTG 120  
Qy 121 MWRRDGACRSRSRQRSKYELACKSNLAHSEPTCVALTFETLVCHHALVYPTL 180  
121 MWRRDGACRSRSRQRSKYELACKSNLAHSEPTCVALTFETLVCHHALVYPTL 180  
Dy 181 PEALQRQMDQEVQDADLITPOGHKLURTFEDAGYLIKTPBENPTQLEGGSIGFE 240  
181 PEALQRQMDQEVQDADLITPOGHKLURTFEDAGYLIKTPBENPTQLEGGSIGFE 240  
Qy 241 TLENCRKAHKELSKEIKRLKGILTQHGIPYTRPTESNLLEHIGHETPRAKSPEQLRGDPG 300  
241 TLENCRKAHKELSKEIKRLKGILTQHGIPYTRPTESNLLEHIGHETPRAKSPEQLRGDPG 300  
Dy 301 LR GSL 305  
Db 301 LR GSL 305  
Dy 301 LR GSL 305  
Db 301 LR GSL 305

RESULT 12

ABW01490  
ID ABW01490 standard; protein; 305 AA.AC ABW01490;  
XX DT 15-JAN-2004 (first entry)

DE Human GlcNAc-phosphotransferase gamma subunit protein.  
XX  
KW Mannose glycoprotein; gene therapy; carbohydrate deficient cell;  
KW lysosomal storage disease; N-acetylglucosamine-1-phosphotransferase;  
KW gastrointestinal; human; enzyme; lectin resistant cell;  
KW deoxymannojirimycin; kifunensine; glycosylation inhibition.  
XX  
OS Homo sapiens.

PA

PI

XX DR

N-PSSB;

ADD27814.

PT

Phosphorylating a protein for treating a patient suffering from a

lysosomal storage disease e.g. Fabry's disease by contacting the protein

with a soluble GlcNAc-phosphotransferase and producing a phosphorylated

protein.

XX PS Claim 11; SEQ ID NO 7; 55pp; English.

CC The invention relates to a method of phosphorylating a protein comprising

CC contacting the protein with a soluble GlcNAc-phosphotransferase (UDP-N-

acetylglucosamine) and producing a phosphorylated protein. The method is

useful for treating a patient suffering from a lysosomal storage disease

e.g. Fabry's disease. The present sequence represents the amino acid

sequence of the human GlcNAc-phosphotransferase gamma subunit.

FT /note= "Mature human GlcNAc-phosphotransferase gamma  
FT subunit protein"  
XX XX  
XX US2003124652-A1.  
PN XX  
FD 03-JUL-2003.  
XX XX  
PP 21-DEC-2-2001; 2001US-00023889.  
XX XX  
PR 21-DEC-2-2001; 2001US-00023889.  
XX PA  
(NOVA-) NOVAZYME PHARM INC.  
XX XX  
PT Canfield NM;  
XX XX  
DR WPI; 2003-810984/76.  
XX N-PSDB; AAB62492.  
XX PR  
PT Producing a high mannose glycoprotein for treating lysosomal storage disease, comprises culturing the lectin resistant mammalian cell in the presence of deoxymannojirimycin and kifunensine.  
PT PT  
XX XX  
PS Claim 9; Page 20; 46pp; English.  
XX DR  
DR  
CC The invention relates to a method for producing a high mannose glycoprotein. The method comprises: introducing and expressing a polynucleotide encoding a glycoprotein into a mammalian cell; culturing the cell in the presence of a lectin to obtain a lectin resistant cell; isolating the cell; culturing the cell in the presence of deoxymannojirimycin and kifunensine to inhibit glycosylation of the glycoprotein; and collecting the glycoprotein. The invention is useful in gene therapy. The method is useful for producing a high mannose glycoprotein in a complex carbohydrate deficient cell for treating lysosomal storage disease. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-phosphotransferase gamma subunit protein Sequence 305 AA;

Query Match 100.0%; Score 1634; DB 7; Length 305;  
Best Local Similarity 100.0%; Pred. No. 3.2e-161; Indels 0; Gaps 0;  
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAGLARILLLGLISAGGPPAPAGAKVVKVTPNATGVNNPLPQASRLQAKRDPSPVG 60  
Db 1 MAAGLARILLLGLISAGGPPAPAGAKVVKVTPNATGVNNPLPQASRLQAKRDPSPVG 60

QY 61 PVLHLFRSLGKCFSLVESTVTKYKEFCPFHVNTOHQFRWNAVSGILGWHWEIANNTFG 120  
Db 61 PVLHLFRSLGKCFSLVESTVTKYKEFCPFHVNTOHQFRWNAVSGILGWHWEIANNTFG 120

QY 121 MMWEDGACRSRSRQSKVELACKSNRLAHVSEPTCVALTPEPVLCHPHALLVPTL 180  
Db 121 MMWEDGACRSRSRQSKVELACKSNRLAHVSEPTCVALTPEPVLCHPHALLVPTL 180

QY 181 PEALQRQDQVEQDADLELITPOGHKRLLRTFEDAGYKIPPEENPTQEGGPSSIGE 240  
Db 181 PEALQRQDQVEQDADLELITPOGHKRLLRTFEDAGYKIPPEENPTQEGGPSSIGE 240

QY 241 TLENCRKAHKELSKERKLGILITQHGIPIPYYRPTETSNLHIGETPRAWSPEQLRGDPG 300  
Db 241 TLENCRKAHKELSKERKLGILITQHGIPIPYYRPTETSNLHIGETPRAWSPEQLRGDPG 300

QY 301 LRSSL 305  
Db 301 LRSSL 305

QY 15-JAN-2004 (first entry)  
XX DT  
DE Human secreted protein #1156.  
XX XX  
KW human secreted protein; Antiallergic; Antiinflammatory; Antibacterial; Anti-HIV; Cytostatic; Immunosuppressive; Hemostatic.  
XX KW  
OS Homo sapiens.  
XX OS  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX XX  
PT PI Rosen CA, Ruben SM;  
XX DR  
DR  
WPI; 2003-140218/13.

XX PR  
PT New human secreted proteins and nucleic acid molecules, useful for preparing a diagnostic or pharmaceutical composition for diagnosing or treating allergic or asthmatic disorders, or related immediate hypersensitivity disorders.  
XX PT  
XX PS  
Claim 1; SEQ ID NO 455; 1233pp; English.

CC The present invention relates to an isolated polypeptide or human secreted protein. The polypeptides, nucleic acid molecules, antibodies or their fragments, and agonists or antagonists that bind are useful for preparing a diagnostic or pharmaceutical composition for diagnosing or treating allergic or asthmatic disorders. The polypeptide is also useful for identifying a binding partner by contacting the polypeptide with a binding partner, and determining whether the binding partner increases or decreases the activity of the polypeptide. The polypeptides and nucleic acid molecules are also useful for detecting, preventing, diagnosing, prognosticating, treating or ameliorating inflammatory disorders, neoplastic diseases, wound healing and disorders of epithelial cell proliferation, immune disorders, cardiovascular disorders, blood-related disorders, infectious diseases, endocrine disorders, or gastrointestinal disorders. The nucleic acids are also useful for chromosome identification, radiation hybrid mapping or long-range restriction mapping, a molecular weight marker, or a hybridization or diagnostic probe. The polypeptides and antibodies are useful for providing immunological probes for differential identification of the tissues secreted protein.  
CC  
XX SQ Sequence 305 AA;

Query Match 100.0%; Score 1634; DB 7; Length 305;  
Best Local Similarity 100.0%; Pred. No. 3.2e-161; Indels 0; Gaps 0;  
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAGLARILLLGLISAGGPPAPAGAKVVKVTPNATGVNNPLPQASRLQAKRDPSPVG 60  
Db 1 MAAGLARILLLGLISAGGPPAPAGAKVVKVTPNATGVNNPLPQASRLQAKRDPSPVG 60

QY 61 PVLHLFRSLGKCFSLVESTVTKYKEFCPFHVNTOHQFRWNAVSGILGWHWEIANNTFG 120  
Db 61 PVLHLFRSLGKCFSLVESTVTKYKEFCPFHVNTOHQFRWNAVSGILGWHWEIANNTFG 120

QY 121 MMWEDGACRSRSRQSKVELACKSNRLAHVSEPTCVALTPEPVLCHPHALLVPTL 180  
Db 121 MMWEDGACRSRSRQSKVELACKSNRLAHVSEPTCVALTPEPVLCHPHALLVPTL 180

QY 181 PEALQRQDQVEQDADLELITPOGHKRLLRTFEDAGYKIPPEENPTQEGGPSSIGE 240  
Db 181 PEALQRQDQVEQDADLELITPOGHKRLLRTFEDAGYKIPPEENPTQEGGPSSIGE 240

Db 181 PEALQRQWQDQEQLADLTLITPOGHEKULRTLFEDAGYLKTPENBPTQLEGGPPDSLIGFE 240  
 QY ||||| TLENCKAHKELSKKEIKLKGLTQHGPYTRPTETSNLEHGHETPRAKSPEQLRGDPG 300  
 Db ||||| 241 TLENCKAHKELSKKEIKLKGLTQHGPYTRPTETSNLEHGHETPRAKSPEQLRGDPG 300  
 QY |||||  
 Db 301 LRGSL 305  
 QY |||||  
 Db 301 LRGSL 305

RESULT 14  
 ADD3804 ADD3804 standard; protein: 305 AA.  
 ID ADD3804;  
 XX AC ADD3804;  
 XX DT 15-JAN-2004 (first entry)  
 XX DE Human secreted protein #267.  
 XX KW human secreted protein; Antiallergic; Antiinflammatory; Antibacterial; Anti-HIV; Cytostatic; Immunosuppressive; Hemostatic.  
 XX OS Homo sapiens.  
 XX PN WO20020526 A2.  
 XX PD 14-NOV-2002.  
 XX PF 19-MAR-2002; 2002WO-US008279.  
 XX PR 21-MAR-2001; 2001US-027734P.  
 PR 19-JUL-2001; 2001US-030617P.  
 PR 13-NOV-2001; 2001US-033128P.  
 XX PA (HUMA-) HUMAN GENOME SCI INC.  
 XX PI Rosen CA, Ruben SM;  
 XX DR WPI; 2003-140218/13.  
 XX PT New human secreted proteins and nucleic acid molecules, useful for preparing a diagnostic or pharmaceutical composition for diagnosing or treating allergic or asthmatic disorders, or related immediate hypersensitivity disorders.  
 XX PS Claim 1, SEQ ID NO 566; 1323pp; English.

The present invention relates to an isolated polypeptide or human secreted protein. The polypeptides, nucleic acid molecules, antibodies or their fragments, and agonists or antagonists that bind are useful for preparing a diagnostic or pharmaceutical composition for diagnosing or treating allergic or asthmatic disorders. The polypeptide is also useful for identifying a binding partner by contacting the polypeptide with a binding partner, and determining whether the binding partner increases or decreases the activity of the polypeptide. The polypeptides and nucleic acid molecules are also useful for detecting, preventing, diagnosing, prognosticating, treating or ameliorating inflammatory disorders, neoplastic diseases, wound healing and disorders of epithelial cell proliferation, immune disorders, cardiovascular disorders, blood-related disorders, infectious diseases, endocrine disorders, or gastrointestinal disorders. The nucleic acids are also useful for chromosome identification, radiation hybrid mapping or long-range restriction mapping, as molecular weight markers, or as hybridization or diagnostic probes. The polypeptides and antibodies are useful for providing immunological probes for differential identification of the tissues secreted protein. The present sequence represents a human sequence 305 AA;

Sequence 305 AA;  
 SQ Query Match. 100.0%; Score 1634; DB 7; Length 305;

---

Best Local Similarity 100.0%; Pred. No. 3.2e-161; Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0; QY 1 MAAGILRLLLGISAGGAPAGAKMVKVEPNAGVNNPLPQASRLQAKRDPSVSG 60  
 Db 1 MAAGILRLLLGISAGGAPAGAKMVKVEPNAGVNNPLPQASRLQAKRDPSVSG 60  
 QY 61 PVHLFLRSKCFSLVESTVYKVECPFHVTQHQFRNAYVSGIJJWHENIANWFTG 120  
 Db 61 PVHLFLRSKCFSLVESTVYKVECPFHVTQHQFRNAYVSGIJJWHENIANWFTG 120  
 QY 121 MWMDGDACRSRSRSQSKYBLACKSRNLAHVSEPTCVYALIFETPVCHPHALLYPTL 180  
 Db 121 MWMDGDACRSRSRSQSKYBLACKSRNLAHVSEPTCVYALIFETPVCHPHALLYPTL 180  
 QY 181 PEALQRQWQDQEQLADLTLITPOGHEKULRTLFEDAGYLKTPENBPTQLEGGPPDSLIGFE 240  
 Db 181 PEALQRQWQDQEQLADLTLITPOGHEKULRTLFEDAGYLKTPENBPTQLEGGPPDSLIGFE 240  
 DE Human GICNAC-phosphotransferase gamma subunit protein.  
 XX KW Glycoprotein; lectin; lysosomal storage disease; gastrointestinal N-acetylglucosamine-1-phosphotransferase; gene therapy; human; enzyme.  
 XX OS Homo sapiens.  
 XX PT Key Location/Qualifiers  
 PT Peptide 1..24  
 PT /label= Signal\_peptide  
 PT Protein 25..305  
 PT /note= "Mature human GICNAC-phosphotransferase gamma subunit protein"  
 XX PN US2003124653-A1.  
 XX PD 03-JUL-2003.  
 XX PF 21-DEC-2001; 2001US-00023890.  
 XX PR 21-DEC-2001; 2001US-00023890.  
 XX PA (NOVA-) NOVAZYME PHARM INC.  
 XX PI Canfield WM;  
 XX DR WPI; 2003-810985/76.  
 XX DR N\_PDB; R082651.  
 XX PT Producing a glycoprotein with reduced complex carbohydrates by culturing the lectin resistant mammalian cell expressing the glycoprotein for treating lysosomal storage disease.  
 XX PS Claim 9, Page 20; 46pp; English.  
 CC The present invention provides a method of producing a glycoprotein having reduced complex carbohydrates by culturing the lectin resistant

CC mammalian cell expressing the glycoprotein. The method is useful for  
 CC producing a glycoprotein with reduced complex carbohydrates for treating  
 CC lysosomal storage disease. The present invention is also useful in gene  
 CC therapy. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-  
 XX phosphotransferase gamma subunit precursor cDNA.

Sequence 305 AA;

Query Match	100 0%	Score 1634;	DB 7;	Length 305;
Best Local Similarity	100.0%	Pred. No. 3.2e-161;		
Matches	305;	Mismatches	0;	
		Indels	0;	Gaps 0;
Qy	1	MAAGLARLLILGLSAGGAPAGAAGKVKTEEPNARGVNPPLPQASRLOAKRKPSPVSG	60	
Db	1	MAAGLARLLILGLSAGGAPAGAAGKVKTEEPNARGVNPPLPQASRLOAKRKPSPVSG	60	
Qy	61	PVHLFRISGKGFLSLVSTYKIEFCPPHNVTQHEQFRWNAYSGIGIWHWEIANNTFG	120	
Qy	61	PVHLFRISGKGFLSLVSTYKIEFCPPHNVTQHEQFRWNAYSGIGIWHWEIANNTFG	120	
Db	121	MWNRDGAACRSRSRQSKEVLAAGKSRLARYSERPSCTCVIALTPEPBLVCPHALLVPTI	180	
Db	121	MWNRDGAACRSRSRQSKEVLAAGKSRLARYSERPSCTCVIALTPEPBLVCPHALLVPTI	180	
Qy	181	PEALQROWDQEQQDALELTTPQGHKEKLRTLPEDAGYLKPEEMPTOEGGPDSLGFE	240	
Qy	181	PEALQROWDQEQQDALELTTPQGHKEKLRTLPEDAGYLKPEEMPTOEGGPDSLGFE	240	
Db	241	TLENCRKAHKELSKETKRLKGILTQHGIPYRPTTSNLIGHTRPASPEOLRGDG	300	
Db	241	TLENCRKAHKELSKETKRLKGILTQHGIPYRPTTSNLIGHTRPASPEOLRGDG	300	
Qy	301	LRGSL 305		
Db	301	LRGSL 305		

Search completed: July 26, 2004, 11:09:12  
 Job time : 30.1358 secs

This Page Blank (uspto)



RESULT 2  
US-09-636-077A-3  
; Sequence 3, Application US/09636077A  
; Patent No. 6537785  
; GENERAL INFORMATION:  
; APPLICANT: CAMPFIELD, WILLIAM M  
; TITLE OF INVENTION: METHODS OF TREATING LYSOSMAL STORAGE DISEASE  
; FILE REFERENCE: 19612US0  
; CURRENT APPLICATION NUMBER: US/09/636,077A  
; CURRENT FILING DATE: 2000-08-10  
; PRIORITY APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 3  
; LENGTH: 305  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: (1)..(24)  
; US-09-636-077A-3

Query Match 100.0%; Score 1634; DB 4; Length 305;  
Best Local Similarity 100.0%; Pred. No. 5.6e-178; Mismatches 0; Indels 0; Gaps 0;  
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAGLARLILGLISAGGAPAGAACOMKVEBNAFGVNNPFLPOASRLQAKRDRSPVSG 60  
Db 1 MAAGLARLILGLISAGGAPAGAACOMKVEBNAFGVNNPFLPOASRLQAKRDRSPVSG 60  
QY 61 PVHFRFLSGKCFSLVESTYKYPECPFHNTQHOTFRWNAASGLGIWHENELIANTFTG 120  
Db 61 PVHFRFLSGKCFSLVESTYKYPECPFHNTQHOTFRWNAASGLGIWHENELIANTFTG 120  
QY 121 MWNRDGDACRSRSRQSKVELACKSNRLAHVSEPTCVYALTFETPLVCHPHALLYPTL 180  
Db 121 MWNRDGDACRSRSRQSKVELACKSNRLAHVSEPTCVYALTFETPLVCHPHALLYPTL 180  
QY 1 MAAGLARLILGLISAGGAPAGAACOMKVEBNAFGVNNPFLPOASRLQAKRDRSPVSG 60  
Db 1 MAAGLARLILGLISAGGAPAGAACOMKVEBNAFGVNNPFLPOASRLQAKRDRSPVSG 60  
QY 61 PVHFRFLSGKCFSLVESTYKYPECPFHNTQHOTFRWNAASGLGIWHENELIANTFTG 120  
Db 61 PVHFRFLSGKCFSLVESTYKYPECPFHNTQHOTFRWNAASGLGIWHENELIANTFTG 120  
QY 121 MWNRDGDACRSRSRQSKVELACKSNRLAHVSEPTCVYALTFETPLVCHPHALLYPTL 180  
Db 121 MWNRDGDACRSRSRQSKVELACKSNRLAHVSEPTCVYALTFETPLVCHPHALLYPTL 180  
QY 181 PEALQRONDQVEQDADLITPQGHKEKLRTFEDAGLYKTPPEENPTQLEGGPSIGFE 240  
Db 181 PEALQRONDQVEQDADLITPQGHKEKLRTFEDAGLYKTPPEENPTQLEGGPSIGFE 240  
QY 241 TLENCRKAHKELSKERIKLKGILTQHGPYTRPTESNLIGHETPRAKSPQEQLRGDPG 300  
Db 241 TLENCRKAHKELSKERIKLKGILTQHGPYTRPTESNLIGHETPRAKSPQEQLRGDPG 300  
QY 301 LRGSIL 305  
Db 301 LRGSIL 305

---

RESULT 4  
US-09-986-552-3  
; Sequence 3, Application US/09986552  
; Patent No. 6670165  
; GENERAL INFORMATION:  
; APPLICANT: CAMPFIELD, William M  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSMAL HYDROLASES  
; FILE REFERENCE: 21500US77DIV  
; CURRENT APPLICATION NUMBER: US/09/986, 552  
; CURRENT FILING DATE: 2001-11-09  
; PRIORITY APPLICATION NUMBER: 09/635,872  
; PRIOR FILING DATE: 2000-08-10  
; PRIORITY APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 3  
; LENGTH: 305  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: (1)..(24)  
; OTHER INFORMATION:  
; US-09-986-552-3

Query Match 100.0%; Score 1634; DB 4; Length 305;  
Best Local Similarity 100.0%; Pred. No. 5.6e-178; Mismatches 0; Indels 0; Gaps 0;  
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAGLARLILGLISAGGAPAGAACOMKVEBNAFGVNNPFLPOASRLQAKRDRSPVSG 60  
Db 1 MAAGLARLILGLISAGGAPAGAACOMKVEBNAFGVNNPFLPOASRLQAKRDRSPVSG 60  
QY 61 PVHFRFLSGKCFSLVESTYKYPECPFHNTQHOTFRWNAASGLGIWHENELIANTFTG 120  
Db 61 PVHFRFLSGKCFSLVESTYKYPECPFHNTQHOTFRWNAASGLGIWHENELIANTFTG 120  
QY 121 MWNRDGDACRSRSRQSKVELACKSNRLAHVSEPTCVYALTFETPLVCHPHALLYPTL 180





TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
 FILE REFERENCE: 195611US0  
 CURRENT APPLICATION NUMBER: US/09/635,872A  
 CURRENT FILING DATE: 2000-08-10  
 PRIOR FILING DATE: 1999-09-14  
 NUMBER OF SEQ ID NOS: 52  
 SOFTWARE: PatentIn version 3.0  
 SEQ ID NO: 28  
 LENGTH: 21  
 TYPE: PRT  
 ORGANISM: Bos taurus  
 US-09-635-872A-28  
 Query Match 6.4%; Score 104; DB 4; Length 21;  
 Best Local Similarity 90.5%; Pred. No. 4.8e-05; Mismatches 1; Indels 0; Gaps 0;  
 Matches 19; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 25 AKMKVVEEPNPFGLNNPFLPQ 45  
 Db 1 AKMKVVEEPNPFGLNNPFLPQ 21  
 RESULT 12  
 US-09-636-077A-28  
 Sequence 28, Application US/09636077A  
 ; Patent No. 6537785  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CANFIELD, WILLIAM  
 ; TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE  
 ; FILE REFERENCE: 195612US0  
 ; CURRENT APPLICATION NUMBER: US/09/635,077A  
 ; CURRENT FILING DATE: 2000-08-10  
 ; PRIOR APPLICATION NUMBER: 60/153,831  
 ; PRIOR FILING DATE: 1999-09-14  
 ; NUMBER OF SEQ ID NOS: 52  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO: 28  
 ; LENGTH: 21  
 ; TYPE: PRT  
 ; ORGANISM: Bos taurus  
 US-09-636-077A-28  
 Query Match 6.4%; Score 104; DB 4; Length 21;  
 Best Local Similarity 90.5%; Pred. No. 4.8e-05; Mismatches 1; Indels 0; Gaps 0;  
 Matches 19; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 25 AKMKVVEEPNPFGLNNPFLPQ 45  
 Db 1 AKMKVVEEPNPFGLNNPFLPQ 21  
 RESULT 13  
 US-09-636-060C-28  
 Sequence 28, Application US/09636060C  
 ; Patent No. 6642038  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CANFIELD, WILLIAM M  
 ; APPLICANT: CANFIELD, WILLIAM M  
 ; APPLICANT: GICNAC, PHOTOPHOTTRANSFERASE OF THE LYSOSOMAL TARGETTING PATHWAY  
 ; FILE REFERENCE: 210119US00CNT  
 ; CURRENT APPLICATION NUMBER: US/09/636,060C  
 ; CURRENT FILING DATE: 2000-08-10  
 ; PRIOR APPLICATION NUMBER: 60/153,831  
 ; PRIOR FILING DATE: 1999-09-14  
 ; NUMBER OF SEQ ID NOS: 52  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO: 28  
 ; LENGTH: 21  
 ; TYPE: PRT  
 ; ORGANISM: Bos taurus  
 US-09-636-077A-28  
 Query Match 6.4%; Score 104; DB 4; Length 21;  
 Best Local Similarity 90.5%; Pred. No. 4.8e-05; Mismatches 1; Indels 0; Gaps 0;  
 Matches 19; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 25 AKMKVVEEPNPFGLNNPFLPQ 45  
 Db 1 AKMKVVEEPNPFGLNNPFLPQ 21  
 RESULT 14  
 US-09-986-552-28  
 Sequence 28, Application US/09986552  
 ; Patent No. 6670165  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CANFIELD, William  
 ; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
 ; FILE REFERENCE: 215083US771IV  
 ; CURRENT APPLICATION NUMBER: US/09/986,552  
 ; CURRENT FILING DATE: 2001-11-09  
 ; PRIOR APPLICATION NUMBER: 09/635,872  
 ; PRIOR FILING DATE: 2000-08-10  
 ; PRIOR APPLICATION NUMBER: 60/153,831  
 ; PRIOR FILING DATE: 1999-09-14  
 ; NUMBER OF SEQ ID NOS: 52  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO: 28  
 ; LENGTH: 21  
 ; TYPE: PRT  
 ; ORGANISM: Bos taurus  
 US-09-986-552-28  
 Query Match 6.4%; Score 104; DB 4; Length 21;  
 Best Local Similarity 90.5%; Pred. No. 4.8e-05; Mismatches 1; Indels 0; Gaps 0;  
 Matches 19; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 25 AKMKVVEEPNPFGLNNPFLPQ 45  
 Db 1 AKMKVVEEPNPFGLNNPFLPQ 21  
 RESULT 15  
 US-09-199-637A-367  
 Sequence 367, Application US/09199637A  
 ; Patent No. 6355411  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ausubel, Frederick  
 ; APPLICANT: Goodman, Howard M.  
 ; APPLICANT: Rahme, Laurence G.  
 ; APPLICANT: Mahajan-Mikkil, Shalina  
 ; APPLICANT: Tan, Man-Wah  
 ; APPLICANT: Cao, Hui  
 ; APPLICANT: Dreher, Eliana  
 ; APPLICANT: Tsongalis, John  
 ; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID  
 ; TITLE OF INVENTION: SEQUENCES AND USES THEREOF  
 ; FILE REFERENCE: 00716/361002  
 ; CURRENT APPLICATION NUMBER: US/09/199,637A  
 ; CURRENT FILING DATE: 1998-11-25  
 ; PRIOR APPLICATION NUMBER: 60/066,517  
 ; PRIOR FILING DATE: 1997-11-25  
 ; NUMBER OF SEQ ID NOS: 437  
 ; SOFTWARE: FastSBQ for Windows Version 4.0  
 ; SEQ ID NO: 367  
 ; LENGTH: 378  
 ; TYPE: PRT  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-199-637A-367  
 Query Match 5.7%; Score 93.5; DB 4; Length 378;  
 Best Local Similarity 20.8%; Pred. No. 0.082; Mismatches 76; Indels 113; Gaps 11;

Do 145 GAGFORAVYMSGVANVRIPRLARSRSQQ ..... 175  
Qy 83 FCPHNVTQHEOPRWNAYSGLIGIWHETIANNTFGMMWRDGD----ACRSRSRSQS 136  
Do 176 ---FANLILHEOT-----EVIG---.WLRDGDRVVGAVATSRGBIRGD 210  
Qy 137 KVBLACGK-SNLAHNSPSTCVYALTFETPYCHPHALLUYPTFELBQMDQVEOL 195  
Do 211 KVBLAAGAWSGEL-----LKPGIABLPVPUVKOMILYKAADFLPR----- 252  
Qy 196 ADELITPOCH-----EKUJTLFEDAGYLKTPEENEPTOLEGGPDLSGFETLENCRK 247  
Do 253 --MVLAGGRYATPRDCHLIGSTHSGEKTBD-----EAESLRA 294  
Qy 248 AHKBLSKTKRKL-----GLL--TONGIPYTRP 273  
Db 295 SAABELPLADMPVAHAGURPGSPEGIPYGP 328

Search completed: July 26, 2004, 11:13:20  
Job time : 9.59705 secs

OM protein - protein search, using sw model  
 Copyright (c) 1993 - 2004 Compugen Ltd.  
 GenCore version 5.1.6

Run on: July 26, 2004, 11:11:44 ; Search time 23.4455 Seconds  
 Sequence: (without alignments)  
 4073.614 Million cell updates/sec

Title: US-10-657-280-3  
 Perfect score: 1624  
 Sequence: MAGLARILLILGLISAGGPA.....TPRAXSPBQLRQDPGLRGSL 305  
 Scoring table: BLOSSUM62  
 Gappp 10.0 , Gapext 0.5

Searched: 1288442 seqs, 313154207 residues  
 Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000  
 Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : Published Applications AA:  
 1: /cgn2\_6/pctodata/1/pubpaal/us07\_PUBCOMB.pep:  
 2: /cgn2\_6/podata/1/pubpaal/PCT\_NEW\_PUB\_pep:  
 3: /cgn2\_6/podata/1/pubpaal/us06\_NEW\_PUB\_pep:  
 4: /cgn2\_6/podata/1/pubpaal/us06\_PUBCOMB\_pep:  
 5: /cgn2\_6/podata/1/pubpaal/us07\_NEW\_PUB\_pep:  
 6: /cgn2\_6/podata/1/pubpaal/PCTUS\_PUBCOMB\_pep:  
 7: /cgn2\_6/podata/1/pubpaal/us08\_NEW\_PUB\_pep:  
 8: /cgn2\_6/podata/1/pubpaal/us08\_PUBCOMB\_pep:  
 9: /cgn2\_6/podata/1/pubpaal/us09A\_PUBCOMB\_pep:  
 10: /cgn2\_6/podata/1/pubpaal/us09C\_PUBCOMB\_pep:  
 11: /cgn2\_6/podata/1/pubpaal/us09C\_PUBCOMB\_pep:  
 12: /cgn2\_6/podata/1/pubpaal/us09\_NEW\_PUB\_pep:  
 13: /cgn2\_6/podata/1/pubpaal/us10\_PUBCOMB\_pep:  
 14: /cgn2\_6/podata/1/pubpaal/us10\_PUBCOMB\_pep:  
 15: /cgn2\_6/podata/1/pubpaal/us10C\_PUBCOMB\_pep:  
 16: /cgn2\_6/podata/1/pubpaal/us10\_NEW\_PUB\_pep:  
 17: /cgn2\_6/podata/1/pubpaal/us10c\_PUBCOMB\_pep:  
 18: /cgn2\_6/podata/1/pubpaal/us60\_PUBCOMB\_pep:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1634	100.0	305	9 US-09-895-072-3	Sequence 3, Appli
2	1634	100.0	305	9 US-09-986-522-3	Sequence 3, Appli
3	1634	100.0	305	14 US-10-023-888-7	Sequence 7, Appli
4	1634	100.0	305	14 US-10-023-889-7	Sequence 7, Appli
5	1634	100.0	305	14 US-10-023-890-7	Sequence 7, Appli
6	1634	100.0	305	14 US-10-024-197-7	Sequence 7, Appli
7	1634	100.0	305	14 US-10-023-894-7	Sequence 7, Appli
8	1634	100.0	305	14 US-10-306-686-3	Sequence 3, Appli
9	1634	100.0	306	12 US-10-372-876-152	Sequence 152, Appli
10	1634	100.0	306	14 US-10-097-065-52	Sequence 152, Appli
11	1634	100.0	310	12 US-10-266-115-1319	Sequence 1319, Appli
12	1288.5	78.9	307	9 US-09-895-072-9	Sequence 9, Appli
13	1288.5	78.9	307	9 US-09-986-522-9	Sequence 9, Appli
14	1288.5	78.9	307	14 US-10-023-888-12	Sequence 12, Appli
15	1288.5	78.9	307	14 US-10-023-889-12	Sequence 12, Appli

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

No.	Score	Quality	Match	Length	DB	ID	Description
-----	-------	---------	-------	--------	----	----	-------------

QY 121 MWMDGACRSRSRQSKVELACKSNSRLAHSPESTCVYALTFETPLVCHALLVPTL 180  
Db 121 MWMDGACRSRSRQSKVELACKSNSRLAHSPESTCVYALTFETPLVCHALLVPTL 180  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: SOLUBLE GICNAC PHOSPHOTRANSFERASE  
; FILE REFERENCE: 20315U577  
; CURRENT APPLICATION NUMBER: US/10/023, 888  
; CURRENT FILING DATE: 2001-12-21  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO: 7  
; LENGTH: 305  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-10-023-888-7

RESULT 2  
US-09-986-552-3  
; Sequence 3, Application US/09986552  
; Patent No. US20020150981A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSSOMAL HYDROLASES  
; FILE REFERENCE: 2150890577DTV  
; CURRENT APPLICATION NUMBER: US/09/986, 552  
; CURRENT FILING DATE: 2001-11-09  
; PRIORITY NUMBER: 09/1635, 872  
; PRIORITY FILING DATE: 2000-08-10  
; PRIORITY APPLICATION NUMBER: 60/153, 831  
; PRIORITY FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO: 3  
; LENGTH: 305  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: (1)..(24)  
; OTHER INFORMATION:  
; US-09-986-552-3

Query Match 100.0%; Score 1634; DB 14; Length 305;  
Best Local Similarity 100.0%; Pred. No. 1..3e-156; Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGALARILLILGLSAGGAPAGAAKMKVVEEPNAFGVNNPFLPOASRLQAKRDPSPVG 60  
Db 1 MAGALARILLILGLSAGGAPAGAAKMKVVEEPNAFGVNNPFLPOASRLQAKRDPSPVG 60  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSSOMAL HYDROLASES  
; FILE REFERENCE: 2150890577DTV  
; CURRENT APPLICATION NUMBER: US/09/986, 552  
; CURRENT FILING DATE: 2001-11-09  
; PRIORITY NUMBER: 09/1635, 872  
; PRIORITY FILING DATE: 2000-08-10  
; PRIORITY APPLICATION NUMBER: 60/153, 831  
; PRIORITY FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO: 3  
; LENGTH: 305  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: (1)..(24)  
; OTHER INFORMATION:  
; US-09-986-552-3

Query Match 100.0%; Score 1634; DB 9; Length 305;  
Best Local Similarity 100.0%; Pred. No. 1..3e-156; Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGALARILLILGLSAGGAPAGAAKMKVVEEPNAFGVNNPFLPOASRLQAKRDPSPVG 60  
Db 1 MAGALARILLILGLSAGGAPAGAAKMKVVEEPNAFGVNNPFLPOASRLQAKRDPSPVG 60  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSSOMAL HYDROLASES  
; FILE REFERENCE: 2150890577DTV  
; CURRENT APPLICATION NUMBER: US/09/986, 552  
; CURRENT FILING DATE: 2001-11-09  
; PRIORITY NUMBER: 09/1635, 872  
; PRIORITY FILING DATE: 2000-08-10  
; PRIORITY APPLICATION NUMBER: 60/153, 831  
; PRIORITY FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO: 3  
; LENGTH: 305  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: (1)..(24)  
; OTHER INFORMATION:  
; US-09-986-552-3

Query Match 100.0%; Score 1634; DB 14; Length 305;  
Best Local Similarity 100.0%; Pred. No. 1..3e-156; Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGALARILLILGLSAGGAPAGAAKMKVVEEPNAFGVNNPFLPOASRLQAKRDPSPVG 60  
Db 1 MAGALARILLILGLSAGGAPAGAAKMKVVEEPNAFGVNNPFLPOASRLQAKRDPSPVG 60  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGH MANNOSE GLYCOPROTEINS IN COMPLI-  
; FILE REFERENCE: 2031120577  
; CURRENT APPLICATION NUMBER: US/10/023, 889  
; CURRENT FILING DATE: 2001-12-21  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO: 7  
; LENGTH: 305  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-10-023-889-7

Query Match 100.0%; Score 1634; DB 14; Length 305;  
Best Local Similarity 100.0%; Pred. No. 1..3e-156; Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGALARILLILGLSAGGAPAGAAKMKVVEEPNAFGVNNPFLPOASRLQAKRDPSPVG 60  
Db 1 MAGALARILLILGLSAGGAPAGAAKMKVVEEPNAFGVNNPFLPOASRLQAKRDPSPVG 60  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGH MANNOSE GLYCOPROTEINS IN COMPLI-  
; FILE REFERENCE: 2031120577  
; CURRENT APPLICATION NUMBER: US/10/023, 889  
; CURRENT FILING DATE: 2001-12-21  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO: 7  
; LENGTH: 305  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-10-023-889-7

Qy 61 PVLFLRISGKCFSLVESTYKVECPFHNTQHEQTRWNAYSGILGIGWHWEIANNTFG 120 ; GENERAL INFORMATION:  
Db 61 PVLFLRISGKCFSLVESTYKVECPFHNTQHEQTRWNAYSGILGIGWHWEIANNTFG 120 ; APPLICANT: CANFIELD, William  
Qy 121 MNMRDGACRSRSRQSKVELAGKSNRLAHVSPESTCVALTPEPFLVCHPHALLVPTL 180 ; TITLE OF INVENTION: HIGHLY PHOSPHORYLATED ACID BETA-GLUCOCEREBROSIDASE AND METHODS  
Db 121 MNMRDGACRSRSRQSKVELAGKSNRLAHVSPESTCVALTPEPFLVCHPHALLVPTL 180 ; FILE REFERENCE: 209794US0  
Qy 181 PEALQRQMDQVEQDIADELTPOGHENKLRTFEDAGSYLKTFEEENETQLEGGPSLGE 240 ; CURRENT APPLICATION NUMBER: US/10/024,197  
Db 181 PEALQRQMDQVEQDIADELTPOGHENKLRTFEDAGSYLKTFEEENETQLEGGPSLGE 240 ; CURRENT FILING DATE: 2001-12-21  
Qy 241 TLENCRKAHKELSKERIKRKGULTOHGI PYTRPTETSNLLEHGHETRAKSPEOLRGDPG 300 ; NUMBER OF SEQ ID NOS: 27  
Db 241 TLENCRKAHKELSKERIKRKGULTOHGI PYTRPTETSNLLEHGHETRAKSPEOLRGDPG 300 ; SOFTWARE: Patentin version 3.1  
Qy 301 LRGSL 305 ; SEQ ID NO: 7  
Db 301 LRGSL 305 ; LENGTH: 305  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-023-890-7  
RESULT 5  
US-10-023-890-7 Application US/10023890  
; Sequence 7, Application No. US20030124653A1  
; Publication No. US20030124653A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: METHOD OF PRODUCING GLYCOPROTEINS HAVING REDUCED COMPLEX CARBOHYDRATE SUBUNITS  
; TITLE OF INVENTION: MAMMALIAN CELLS  
; FILE REFERENCE: 203510577  
; CURRENT APPLICATION NUMBER: US/10/023, 890  
; CURRENT FILING DATE: 2001-12-21  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO: 7  
; LENGTH: 305  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-023-890-7  
Query Match 100.0%; Score 1634; DB 14; Length 305;  
Best Local Similarity 100.0%; Pred. No. 1. 3e-156; Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MAGLARILLILGLSAGGPPAGAAKMKVVEPNAGVNMPFLPQSRLOAKRDPSPVG 60 ; Query Match 100.0%; Score 1634; DB 14; Length 305;  
Db 1 MAGLARILLILGLSAGGPPAGAAKMKVVEPNAGVNMPFLPQSRLOAKRDPSPVG 60 ; Best Local Similarity 100.0%; Pred. No. 1. 3e-156; Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 61 PVLFLRISGKCFSLVESTYKVECPFHNTQHEQTRWNAYSGILGIGWHWEIANNTFG 120 ; Query Match 100.0%; Score 1634; DB 14; Length 305;  
Db 61 PVLFLRISGKCFSLVESTYKVECPFHNTQHEQTRWNAYSGILGIGWHWEIANNTFG 120 ; Best Local Similarity 100.0%; Pred. No. 1. 3e-156; Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 121 MNMRDGACRSRSRQSKVELAGKSNRLAHVSPESTCVALTPEPFLVCHPHALLVPTL 180 ; Query Match 100.0%; Score 1634; DB 14; Length 305;  
Db 121 MNMRDGACRSRSRQSKVELAGKSNRLAHVSPESTCVALTPEPFLVCHPHALLVPTL 180 ; Best Local Similarity 100.0%; Pred. No. 1. 3e-156; Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 181 PEALQRQMDQVEQDIADELTPOGHENKLRTFEDAGSYLKTFEEENETQLEGGPSLGE 240 ; Query Match 100.0%; Score 1634; DB 14; Length 305;  
Db 181 PEALQRQMDQVEQDIADELTPOGHENKLRTFEDAGSYLKTFEEENETQLEGGPSLGE 240 ; Best Local Similarity 100.0%; Pred. No. 1. 3e-156; Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 241 TLENCRKAHKELSKERIKRKGULTOHGI PYTRPTETSNLLEHGHETRAKSPEOLRGDPG 300 ; Query Match 100.0%; Score 1634; DB 14; Length 305;  
Db 241 TLENCRKAHKELSKERIKRKGULTOHGI PYTRPTETSNLLEHGHETRAKSPEOLRGDPG 300 ; Best Local Similarity 100.0%; Pred. No. 1. 3e-156; Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 301 LRGSL 305 ; Query Match 100.0%; Score 1634; DB 14; Length 305;  
Db 301 LRGSL 305 ; Best Local Similarity 100.0%; Pred. No. 1. 3e-156; Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
US-10-023-894-7  
RESULT 7  
US-10-023-894-7 Application US/10023894  
; Sequence 7, Application No. US/10023894  
; Publication No. US20030143669A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; APPLICANT: KORNFIELD, Stuart  
; TITLE OF INVENTION: EXPRESSION OF LYSSOMAL HYDROLASE IN CELLS EXPRESSING PRO-N-ACETYLGUCOSAMINIDASE  
; FILE REFERENCE: 211130577  
; CURRENT APPLICATION NUMBER: US/10/023, 894  
; CURRENT FILING DATE: 2001-12-21  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO: 7  
; LENGTH: 305  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-023-894-7  
Query Match 100.0%; Score 1634; DB 14; Length 305;  
Best Local Similarity 100.0%; Pred. No. 1. 3e-156; Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MAGLARILLILGLSAGGPPAGAAKMKVVEPNAGVNMPFLPQSRLOAKRDPSPVG 60 ; Query Match 100.0%; Score 1634; DB 14; Length 305;  
Db 1 MAGLARILLILGLSAGGPPAGAAKMKVVEPNAGVNMPFLPQSRLOAKRDPSPVG 60 ; Best Local Similarity 100.0%; Pred. No. 1. 3e-156; Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 61 PVLFLRISGKCFSLVESTYKVECPFHNTQHEQTRWNAYSGILGIGWHWEIANNTFG 120 ; Best Local Similarity 100.0%; Pred. No. 1. 3e-156; Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 61 PVLFLRISGKCFSLVESTYKVECPFHNTQHEQTRWNAYSGILGIGWHWEIANNTFG 120 ; Best Local Similarity 100.0%; Pred. No. 1. 3e-156; Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DO ;  
 51 PVLFLFRSLGKCFELVESTYKIECPFHNTQHOTFRNWAQGIGIWHENIAANTFTG 120  
 QY |||||  
 121 MWHRDGDAACRSRQRSKVELACKSNRLAHVSEPTCVALTPEPFLVCHHALVPTL 180  
 DO ;  
 121 MWHRDGDAACRSRQRSKVELACKSNRLAHVSEPTCVALTPEPFLVCHHALVPTL 180  
 QY ;  
 181 PEALQRQDQVEQDADLETLITPOGHEKLRKTLPEDAGYLKTPENEPTEQEGGPDSLGFE 240  
 DO ;  
 181 PEALQRQDQVEQDADLETLITPOGHEKLRKTLPEDAGYLKTPENEPTEQEGGPDSLGFE 240  
 QY ;  
 241 TLENCRAKAKESKEIKRLKGILTQHGIPTYRPTETSNLEIGHETPRAKSPEQQLRGDPG 300  
 DO ;  
 241 TLENCRAKAKESKEIKRLKGILTQHGIPTYRPTETSNLEIGHETPRAKSPEQQLRGDPG 300  
 QY ;  
 301 LRGSL 305  
 DO ;  
 301 LRGSL 305

RESULT 8  
 US-10-306-616-3  
 ; Sequence 3, Application US/10306686  
 ; Publication No. US2003020471A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CAMFIELD, WILLIAM  
 ; TITLE OF INVENTION: PHODOSTER ALPHA-GLUCONACASE OF THE LYOSOMAL TARGETING PATHWAY  
 ; FILE REFERENCE: 2303971US7D7V  
 ; CURRENT APPLICATION NUMBER: US/10/306,686  
 ; CURRENT FILING DATE: 2002-11-29  
 ; PRIOR APPLICATION NUMBER: 09/1636,596  
 ; PRIOR FILING DATE: 2000-08-10  
 ; PRIOR APPLICATION NUMBER: 60/153,831  
 ; PRIOR FILING DATE: 1999-08-14  
 ; NUMBER OF SEQ ID NOS: 52  
 ; SEQ ID NO 3  
 ; LENGTH: 305  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE: NAME/KEY: SIGNAL  
 ; LOCATION: (1) . (24)  
 ; OTHER INFORMATION:  
 US-10-306-616-3

Query Match 100.0%; Score 1634; DB 14; Length 305;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-156; Indels 0; Gaps 0;  
 Matches 305; Conservative 0; Mismatches 0; Name: Signal

QY 1 MAAGLARILLLGLISAGGPPAPAGAAKMKVVEEPNAFGVNNPLPQASRLQAKRDPSVG 60  
 DO ;  
 1 MAAGLARILLLGLISAGGPPAPAGAAKMKVVEEPNAFGVNNPLPQASRLQAKRDPSVG 60  
 QY 61 PVHLFRLSGKCFSLVESTYKIECPFHNTQHOTFRNWAQGIGIWHENIAANTFTG 120  
 DO ;  
 61 PVHLFRLSGKCFSLVESTYKIECPFHNTQHOTFRNWAQGIGIWHENIAANTFTG 120  
 QY 61 PVHLFRLSGKCFSLVESTYKIECPFHNTQHOTFRNWAQGIGIWHENIAANTFTG 120  
 DO ;  
 61 PVHLFRLSGKCFSLVESTYKIECPFHNTQHOTFRNWAQGIGIWHENIAANTFTG 120  
 QY 121 MWHRDGDAACRSRQRSKVELACKSNRLAHVSEPTCVALTPEPFLVCHHALVPTL 180  
 DO ;  
 121 MWHRDGDAACRSRQRSKVELACKSNRLAHVSEPTCVALTPEPFLVCHHALVPTL 180  
 QY ;  
 181 PEALQRQDQVEQDADLETLITPOGHEKLRKTLPEDAGYLKTPENEPTEQEGGPDSLGFE 240  
 DO ;  
 181 PEALQRQDQVEQDADLETLITPOGHEKLRKTLPEDAGYLKTPENEPTEQEGGPDSLGFE 240  
 QY ;  
 241 TLENCRAKAKESKEIKRLKGILTQHGIPTYRPTETSNLEIGHETPRAKSPEQQLRGDPG 300  
 DO ;  
 241 TLENCRAKAKESKEIKRLKGILTQHGIPTYRPTETSNLEIGHETPRAKSPEQQLRGDPG 300  
 QY ;  
 301 LRGSL 305  
 DO ;  
 301 LRGSL 305

RESULT 9  
 US-10-372-876-152  
 ; Sequence 152, Application US/10372876  
 ; Publication No. US2003020471A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Moore, Paul A. et al.  
 ; TITLE OF INVENTION: 110 Human Secreted Proteins  
 ; FILE REFERENCE: PZ0211  
 ; CURRENT APPLICATION NUMBER: US/10/372,876  
 ; CURRENT FILING DATE: 2003-02-26  
 ; PRIOR APPLICATION NUMBER: 09/134,595  
 ; PRIOR FILING DATE: 1999-06-17  
 ; PRIOR APPLICATION NUMBER: PCT-US98/27059  
 ; PRIOR FILING DATE: 1998-12-17  
 ; PRIOR APPLICATION NUMBER: 60/070,923  
 ; PRIOR FILING DATE: 1997-12-18  
 ; PRIOR APPLICATION NUMBER: 60/068,369  
 ; PRIOR FILING DATE: 1997-12-18  
 ; PRIOR APPLICATION NUMBER: 60/068,007  
 ; PRIOR FILING DATE: 1997-12-18  
 ; PRIOR APPLICATION NUMBER: 60/068,057  
 ; PRIOR FILING DATE: 1997-12-18  
 ; PRIOR APPLICATION NUMBER: 60/068,006  
 ; PRIOR FILING DATE: 1997-12-18  
 ; PRIOR APPLICATION NUMBER: 60/068,368  
 ; PRIOR FILING DATE: 1997-12-19  
 ; PRIOR APPLICATION NUMBER: 60/068,169  
 ; PRIOR FILING DATE: 1997-12-19  
 ; Remaining Prior Application data removed - See File Wrapper or PALM.  
 ; NUMBER OF SEQ ID NOS: 672  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 152  
 ; LENGTH: 306  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: SITE  
 ; LOCATION: (306)  
 ; OTHER INFORMATION: Xaa equals stop translation  
 US-10-372-876-152

Query Match 100.0%; Score 1634; DB 12; Length 306;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-156; Indels 0; Gaps 0;  
 Matches 305; Conservative 0; Mismatches 0; Name: Signal

QY 1 MAAGLARILLLGLISAGGPPAPAGAAKMKVVEEPNAFGVNNPLPQASRLQAKRDPSVG 60  
 DO ;  
 1 MAAGLARILLLGLISAGGPPAPAGAAKMKVVEEPNAFGVNNPLPQASRLQAKRDPSVG 60  
 QY 61 PVHLFRLSGKCFSLVESTYKIECPFHNTQHOTFRNWAQGIGIWHENIAANTFTG 120  
 DO ;  
 61 PVHLFRLSGKCFSLVESTYKIECPFHNTQHOTFRNWAQGIGIWHENIAANTFTG 120  
 QY 121 MWHRDGDAACRSRQRSKVELACKSNRLAHVSEPTCVALTPEPFLVCHHALVPTL 180  
 DO ;  
 121 MWHRDGDAACRSRQRSKVELACKSNRLAHVSEPTCVALTPEPFLVCHHALVPTL 180  
 QY ;  
 181 PEALQRQDQVEQDADLETLITPOGHEKLRKTLPEDAGYLKTPENEPTEQEGGPDSLGFE 240  
 DO ;  
 181 PEALQRQDQVEQDADLETLITPOGHEKLRKTLPEDAGYLKTPENEPTEQEGGPDSLGFE 240  
 QY ;  
 241 TLENCRAKAKESKEIKRLKGILTQHGIPTYRPTETSNLEIGHETPRAKSPEQQLRGDPG 300  
 DO ;  
 241 TLENCRAKAKESKEIKRLKGILTQHGIPTYRPTETSNLEIGHETPRAKSPEQQLRGDPG 300  
 QY ;  
 301 LRGSL 305  
 DO ;  
 301 LRGSL 305

RESULT 10  
 US-10-097-065-152 ; Sequence 152, Application US/10097065  
 Publication No. US20030055236A1  
 GENERAL INFORMATION:  
 ; APPLICANT: Moore, Paul A. et al.  
 ; TITLE OF INVENTION: 110 Human Secreted Proteins  
 FILE REFERENCE: P2021P1  
 CURRENT APPLICATION NUMBER: US/10/097, 065  
 ; CURRENT FILING DATE: 2002-03-14  
 ; PRIOR APPLICATION NUMBER: PCT/US98/27059  
 ; PRIOR FILING DATE: 1998-12-17  
 ; PRIOR APPLICATION NUMBER: 60/070, 923  
 ; PRIOR FILING DATE: 1997-12-18  
 ; PRIOR APPLICATION NUMBER: 60/068, 007  
 ; PRIOR FILING DATE: 1997-12-18  
 ; PRIOR APPLICATION NUMBER: 60/068, 057  
 ; PRIOR FILING DATE: 1997-12-18  
 ; PRIOR APPLICATION NUMBER: 60/068, 006  
 ; PRIOR FILING DATE: 1997-12-18  
 ; PRIOR APPLICATION NUMBER: 60/068, 369  
 ; PRIOR FILING DATE: 1997-12-19  
 ; PRIOR APPLICATION NUMBER: 60/068, 367  
 ; PRIOR FILING DATE: 1997-12-19  
 ; PRIOR APPLICATION NUMBER: 60/068, 368  
 ; PRIOR FILING DATE: 1997-12-19  
 ; PRIOR APPLICATION NUMBER: 60/068, 169  
 ; PRIOR FILING DATE: 1997-12-19  
 ; PRIOR APPLICATION NUMBER: 60/068, 053  
 ; PRIOR FILING DATE: 1997-12-18  
 ; PRIOR APPLICATION NUMBER: 60/068, 064  
 ; PRIOR FILING DATE: 1997-12-18  
 ; PRIOR APPLICATION NUMBER: 60/068, 054  
 ; PRIOR FILING DATE: 1997-12-18  
 ; PRIOR APPLICATION NUMBER: 60/068, 008  
 ; PRIOR FILING DATE: 1997-12-18  
 ; PRIOR APPLICATION NUMBER: 60/068, 365  
 ; PRIOR FILING DATE: 1997-12-19  
 ; NUMBER OF SEQ ID NOS: 672  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 152  
 LENGTH: 310  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: SITE  
 LOCATION: (306)  
 OTHER INFORMATION: Xaa equals stop translation  
 US-10-097-065-152

Query Match 100.0%; Score 1634; DB 14; Length 310;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-156;  
 Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAGLARULLGLISAGGPPAPAGAKMKVMEVNPNAFGVNPNPLPQASRLQAKRDPSPVSG 60  
 Db 6 MAAGLARULLGLISAGGPPAPAGAKMKVMEVNPNAFGVNPNPLPQASRLQAKRDPSPVSG 65  
 Qy 61 PVHFLRSKCSFSLVESTYKYCPEPNTQHQTFWNAVSGILGWHEWIANNTFG 120  
 Db 66 PVHFLRSKCSFSLVESTYKYCPEPNTQHQTFWNAVSGILGWHEWIANNTFG 125

QY 121 MWQRDGACRSRSRQSRYBLACKSNRHLAVEPSTVYALTPETPVCHPHALLVPTL 180  
 Db 126 MWQRDGACRSRSRQSRYBLACKSNRHLAVEPSTVYALTPETPVCHPHALLVPTL 185

QY 181 PEALQRQMDQVEQDIADELTPOGHENKLRTFEDAGYKLTBENEQTOLEGPDSDIGFE 240  
 Db 186 PEALQRQMDQVEQDIADELTPOGHENKLRTFEDAGYKLTBENEQTOLEGPDSDIGFE 245

QY 241 TLENCRKAHKESKEIKRKGILQHJIPYTRPTETNLHGHETRAKSPEQLQDGP 300  
 Db 246 PEALQRQMDQVEQDIADELTPOGHENKLRTFEDAGYKLTBENEQTOLEGPDSDIGFE 305

Query Match 100.0%; Score 1634; DB 14; Length 310;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-156; Mismatches 0; Indels 0; Gaps 0;  
 Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAGLARULLGLISAGGPPAPAGAKMKVMEVNPNAFGVNPNPLPQASRLQAKRDPSPVSG 60  
 Db 1 MAGLARULLGLISAGGPPAPAGAKMKVMEVNPNAFGVNPNPLPQASRLQAKRDPSPVSG 60

Qy 61 PHLFRLSKCSFSLVESTYKYCPEPNTQHQTFWNAVSGILGWHEWIANNTFG 120  
 Db 61 PHLFRLSKCSFSLVESTYKYCPEPNTQHQTFWNAVSGILGWHEWIANNTFG 120

Qy 121 MWRDGACRSRSRQSRYBLACKSNRHLAVEPSTVYALTPETPVCHPHALLVPTL 180  
 Db 121 MWRDGACRSRSRQSRYBLACKSNRHLAVEPSTVYALTPETPVCHPHALLVPTL 180

Qy 181 PEALQRQMDQVEQDIADELTPOGHENKLRTFEDAGYKLTBENEQTOLEGPDSDIGFE 240  
 Db 181 PEALQRQMDQVEQDIADELTPOGHENKLRTFEDAGYKLTBENEQTOLEGPDSDIGFE 240

Qy 241 TLENCRKAHKESKEIKRKGILQHJIPYTRPTETNLHGHETRAKSPEQLQDGP 300  
 Software: PatentIn version 3.1

RESULT 11  
 US-10-236-115-1319 ; Sequence 1319, Application US/10236115  
 Publication No. US20040053248A1  
 GENERAL INFORMATION:  
 ; APPLICANT: HYSEQ Inc  
 ; TITLE OF INVENTION: No. US20040053248A1 eel Nucleic Acids and Polypeptides  
 FILE REFERENCE: 784PCT  
 CURRENT APPLICATION NUMBER: US/10/296, 115  
 ; CURRENT FILING DATE: 2002-11-18  
 ; PRIOR APPLICATION NUMBER: US09/488, 725  
 ; PRIOR FILING DATE: 2000-01-21  
 ; PRIOR APPLICATION NUMBER: US09/552, 317  
 ; PRIOR FILING DATE: 2000-04-25  
 ; SBO ID NO 1319  
 ; NUMBER OF SEQ ID NOS: 1478  
 ; LENGTH: 310  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-236-115-1319

Query Match 100.0%; Score 1634; DB 12; Length 310;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-156;  
 Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAGLARULLGLISAGGPPAPAGAKMKVMEVNPNAFGVNPNPLPQASRLQAKRDPSPVSG 60  
 Db 6 MAAGLARULLGLISAGGPPAPAGAKMKVMEVNPNAFGVNPNPLPQASRLQAKRDPSPVSG 65  
 Qy 61 PVHFLRSKCSFSLVESTYKYCPEPNTQHQTFWNAVSGILGWHEWIANNTFG 120  
 Db 66 PVHFLRSKCSFSLVESTYKYCPEPNTQHQTFWNAVSGILGWHEWIANNTFG 125

QY 121 MWQRDGACRSRSRQSRYBLACKSNRHLAVEPSTVYALTPETPVCHPHALLVPTL 180  
 Db 126 MWQRDGACRSRSRQSRYBLACKSNRHLAVEPSTVYALTPETPVCHPHALLVPTL 185

QY 181 PEALQRQMDQVEQDIADELTPOGHENKLRTFEDAGYKLTBENEQTOLEGPDSDIGFE 240  
 Db 186 PEALQRQMDQVEQDIADELTPOGHENKLRTFEDAGYKLTBENEQTOLEGPDSDIGFE 245

QY 241 TLENCRKAHKESKEIKRKGILQHJIPYTRPTETNLHGHETRAKSPEQLQDGP 300  
 Db 246 TLENCRKAHKESKEIKRKGILQHJIPYTRPTETNLHGHETRAKSPEQLQDGP 305

RESULT 12  
 US-09-835-072-9 ; Sequence 9, Application US/09895072  
 ; Sequence No. US2002005550A1  
 ; PATENT NO. US2002005550A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CANFIELD, WILLIAM M  
 ; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYOSOMAL HYDROLASES  
 FILE REFERENCE: 210119US0001  
 CURRENT APPLICATION NUMBER: US/09/895, 072  
 ; CURRENT FILING DATE: 2001-07-02  
 ; PRIOR APPLICATION NUMBER: 60/153, 831  
 ; PRIOR FILING DATE: 1999-09-14  
 ; PRIOR APPLICATION NUMBER: US 09/335, 872  
 ; PRIOR FILING DATE: 2000-08-10  
 ; NUMBER OF SEQ ID NOS: 52  
 ; SOFTWARE: PatentIn version 3.1

;

SEQ ID NO 9  
LENGTH: 307  
TYPE: PRT  
ORGANISM: Mus musculus  
US-09-895-072-9

Query Match 78.9%; Score 1288.5; DB 9; Length 307;  
Best Local Similarity 78.4%; Pred. No. 1.3e-121; Mismatches 40; Indels 1; Gaps 1;  
Matches 240; Conservative 25; Mismatches 40; Indels 1; Gaps 1;

QY 1 MAAGLARILLLGIGSAGGAPAGAAKQRKVERPNAFGVNNPLPQASRLOAKRDPSPVSG 60  
DB 1 MAGLAGTMLGLASQCPAPACAGKQKVVEPTNGNINPLPQASRLOPKRESEAVSG 60

QY 61 PVHLFRSLGSKCISLVESTYKYPFCPPNTQHEQTRWNAVSGIGIWHMELIANFTG 120  
DB 61 PVHLFRLAGCISLVESTYKYPFCPPNTQHEQTRWNAVSGIGIWHMELIANFTG 120

QY 121 MWYRDGDCRSRSRQSKYELACKNSRNRAHVBSPSTVYALJFETPLVCHHALVYPTL 180  
DB 121 MWYRDGDCHSRSRQSKYELACKNSRNRAHVBSPSTVYALJFETPLVCHHALVYPTL 180

QY 181 PEALQRQMDQVEQDADLITPOGHKEKURTFEDAGYKTPPEENPTQLEGGPDIGFE 240  
DB 181 SEALQQRDQVEQDADLITPOGHKEKURTFEDAGYKTPPEENPTQLEGGPDIGFE 240

QY 241 TLENCRKAHKESKEIKLGILTQHGPYTRPTETSNLIGHETP-RASKSPECRGDP 299  
DB 241 TLENCRKAHKESKEIKLGILTQHGPYTRPTETSNLIGHETP-RASKSPECRGDP 299

QY 300 GLRGSL 305  
DB 301 GLRGNI 306

RESULT 13  
US-09-886-552-9  
Sequence 9, Application US/09986552  
Patent No. US2002010981A1

GENERAL INFORMATION:  
APPLICANT: CANFIELD, William  
TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSMAL HYDROLASES  
FILE REFERENCE: 210089US77D17  
CURRENT APPLICATION NUMBER: US/09/986, 552  
CURRENT FILING DATE: 2001-11-09  
PRIORITY NUMBER: 09/1635, 872  
PRIORITY FILING DATE: 2000-08-10  
PRIOR APPLICATION NUMBER: 60/153, 831  
NUMBER OF SEQ ID NOS: 52  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 9  
LENGTH: 307  
TYPE: PRT  
ORGANISM: Mus musculus  
US-10-023-888-12  
US-10-023-888-12  
Sequence 12, Application US/10023888  
Publication No. US20030119088A1

GENERAL INFORMATION:  
APPLICANT: CANFIELD, William  
TITLE OF INVENTION: SOLUBLE GLCNAC PHOSPHOTRANSFERASE  
FILE REFERENCE: 20315US77  
CURRENT APPLICATION NUMBER: US/10/023, 888  
CURRENT FILING DATE: 2001-12-21  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 12  
LENGTH: 307  
TYPE: PRT  
ORGANISM: Mus musculus  
US-10-023-888-12

Query Match 78.9%; Score 1288.5; DB 14; Length 307;  
Best Local Similarity 78.4%; Pred. No. 1.3e-121; Mismatches 40; Indels 1; Gaps 1;  
Matches 240; Conservative 25; Mismatches 40; Indels 1; Gaps 1;

QY 1 MAAGLARILLLGIGSAGGAPAGAAKQRKVERPNAFGVNNPLPQASRLOAKRDPSPVSG 60  
DB 1 MAGLAGTMLGLASQCPAPACAGKQKVVEPTNGNINPLPQASRLOPKRESEAVSG 60

QY 61 PVHLFRSLGSKCISLVESTYKYPFCPPNTQHEQTRWNAVSGIGIWHMELIANFTG 120  
DB 61 PVHLFRLAGCISLVESTYKYPFCPPNTQHEQTRWNAVSGIGIWHMELIANFTG 120

QY 121 MWYRDGDCRSRSRQSKYELACKNSRNRAHVBSPSTVYALJFETPLVCHHALVYPTL 180  
DB 121 MWYRDGDCHSRSRQSKYELACKNSRNRAHVBSPSTVYALJFETPLVCHHALVYPTL 180

QY 181 PEALQRQMDQVEQDADLITPOGHKEKURTFEDAGYKTPPEENPTQLEGGPDIGFE 240  
DB 181 SEALQQRDQVEQDADLITPOGHKEKURTFEDAGYKTPPEENPTQLEGGPDIGFE 240

QY 241 TLENCRKAHKESKEIKLGILTQHGPYTRPTETSNLIGHETP-RASKSPECRGDP 299  
DB 241 TLENCRKAHKESKEIKLGILTQHGPYTRPTETSNLIGHETP-RASKSPECRGDP 299

QY 300 GLRGSL 305  
DB 301 GLRGNI 306

RESULT 15  
US-10-1023-888-12  
US-10-1023-888-12  
Sequence 12, Application US/10023889  
Publication No. US20030124652A1

GENERAL INFORMATION:  
APPLICANT: CANFIELD, William  
TITLE OF INVENTION: METHODS OF PRODUCING HIGH MANNOSE GLYCOPROTEINS IN COMPLEX CARBOOXYLIC ACIDS  
FILE REFERENCE: 203512US77  
CURRENT APPLICATION NUMBER: US/10/023, 889  
CURRENT FILING DATE: 2001-12-21  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: PatentIn version 3.1

; SEQ ID NO 12  
; LENGTH: 307  
; TYPE: PRT  
; ORGANISM: *Mus musculus*  
US-10-023-889-12

Query Match 78.9%; Score 1288.5; DB 14; Length 307;  
Best Local Similarity 78.4%; Pred. No. 1.3e-121;  
Matches 240; Conservative 25; Mismatches 40; Indels 1; Gaps 1;

Qy 1 MAGLARLLILGLSAGGPPAGAAGKRVTEEPNAFGVNMPFLPOASRLQAKRDPSPVSG 60  
Db 1 MAGLAGFLMLIGLASQGPAPACAGMKVVEEPNTFLNPNFLPQASRQPKREPSAVG 60  
Qy 61 PHLFLRSKGKFSLVIVSTYKCBEFCPPHNTQHEQFRWAVASGLGIGWHREIAANTFG 120  
Db 61 PHLFLRAGKCFSLVIVSTYKCBEFCPPHNTQHEQFRWAVASGLGIGWHREIAINTFG 120  
Qy 121 MNRRDGDACRRSRSSRKRSPKVELTGGKSRLAIVYSEPSCTCVVLTFFERPLVCHPHALLVPTL 180  
Db 121 MNVIDGDSCHSKRSRKRSPKVELTGGKSRLAIVYSEPSCTCVVLTFFERPLVCHPHALLVPTL 180  
Qy 181 PEAHQWQDQEQPDLADELTPOGHKEKLRTFEDAGYKTPPEEMPTOEGPDLSGE 240  
Db 181 SBAQQRDLQEQPDLADELTPOGHKEKLRTFEDAGYKTPPEEMPTOEGPDLSGE 240  
Qy 241 TLENCRKAHELSKEKRLKSLTQHICIPYRPTESNLHGHPT-RAKSPERLGP 299  
Db 241 TLDCRKAHAELESQEVRQLTSLQHGPQPTPTTHSQRQQLPIGAIAEHLRSDP 300  
Qy 300 GIGSL 305  
Db 301 GIGNT 305

Search completed: July 26, 2004, 11:25:26  
Job time : 23.4465 secs

This Page Blank (uspto)

Copyright (c) 1993 - 2004 Compugen Ltd.

GenCore version 5.1.6

protein - protein search, using sw model

Search time 7.22934 Seconds  
(without alignments)  
4058.242 Million cell updates/sec

date: July 26, 2004, 11:05:38 ;  
in on:

title: US-10-657-280-3  
effect: score: 1634  
sequence: 1 MAAGLARLLILGLSAGGPA.....TPRAKSPEQLRQDPGLRQSL 305  
scoring table: BL05M62  
Gapop 10.0 , Gapext 0.5

total number of hits satisfying chosen parameters: 283366

minimum DB seq length: 0  
maximum DB seq length: 2000000000

post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

database : PIR\_7B:  
1: pir1:  
2: pir2:  
3: pir3:  
4: pir4:  
pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

### SUMMARIES

result No.	Score	Query Match Length	DB ID	Description
1	1634	100.0	T45062	hypothetical protein c316G12.3 [imported] - human
2	185.5	11.4	T24944	hypothetical protein OS-9 protein precursor
3	140.5	8.6	JC2889	80k protein H precursor
4	105.6	6.5	A33469	hypothetical protein
5	93.5	5.7	T01598	hypothetical protein
6	93.5	216	F70318	hypothetical protein
7	93.5	5.7	A48343	24K hypothetical protein
8	93.5	1436	F89904	conserved hypothetical protein
9	92	5.6	T43152	hypothetical protein
10	92	5.6	I36511	involutrin L - dou
11	92	5.6	C59431	KIAA391 protein [hypothetical protein]
12	91.5	5.6	A39926	probable protein kinase
13	91.5	5.6	T41623	major surface glycoprotein
14	91.5	1014	T30545	TMA box binding protein
15	91	5.6	T47587	involutrin L - gorgonin L - gorgonin
16	90.5	5.5	I37060	involutrin L - gorgonin L - gorgonin
17	90	5.5	T03162	NLR protein - human
18	90	5.5	T03454	tetraricopeptidase
19	89	2025	JC020	kyurenine-oxoprolut kinase [imported]
20	88	5.4	S66270	TyB protein - yeast
21	88	5.4	T51930	probable D-amino acid oxidase
22	88	5.4	S52481	involutrin M - gorgonin
23	87.5	5.4	T83078	titin, cardiac muscle
24	87.5	605	I37061	similar to axillin
25	87.5	5.4	I38349	ubiquitin-protein ligase
26	87	638	E84799	involutrin - cotoxin
27	87	5.3	T30812	hypothetical protein
28	86.5	493	A57783	C;Species: Caenorhabditis elegans
29	86	5.3	B86285	hypothetical protein

### ALIGNMENTS

RESULT 1

T45062  
hypothetical protein c316G12.3 [imported] - human  
C;Species: Homo sapiens (man)  
C;Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000  
C;Accession: T45062  
R;Frankland, J.  
submitted to the EMBL Data Library, July 1999  
A;Reference number: Z22901  
A;Accession: T45062  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-305 <FFRA>  
A;Cross-references: EMBL:AU031709; PIDN: CAB56184.1  
A;Experimental source: clone LA16-316G12  
C;Genetics:  
A;Introns: 18/1; 37/2; 60/1; 78/2; 106/2; 137/3; 176/1; 203/3; 247/3; 275/1  
A;Note: c316G12.3

Query Match Score 100.0%; DB 2; length 305; Best Local Similarity 100.0%; Pred. No. 3e-132; Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0; gaps 0; Query QY Description

1 MAAGLARLLILGLSAGGPAAGAKVVEENAFGVNNPFLPQARLOAKRDPSVG 60  
Db 1 MAAGLARLLILGLSAGGPAAGAKMVKVEEPNAFGVNPNFLPQASRLQAKRDPSPVSG 60

QY 61 PVHHLRSKCFPSLWESTYKVECPFHNTQHQFRWNNAYSGILGIVHEWEIANTFTG 120  
Db 61 PVHHLRSKCFPSLWESTYKVECPFHNTQHQFRWNNAYSGILGIVHEWEIANTFTG 120

QY 121 MMWRDGAERSRSTQSKVTLACGSKSNLAHVSAPSCTVIALFPEPLVCHPHALVYPTL 180  
Db 121 MMWRDGAERSRSTQSKVTLACGSKSNLAHVSAPSCTVIALFPEPLVCHPHALVYPTL 180

QY 181 PEALQRQWDOVEQDADELTIPQHCKLRTLFEDAGVLUKTPEENEPITOLEGPGPDLSGF 240  
Db 181 PEALQRQWDOVEQDADELTIPQHCKLRTLFEDAGVLUKTPEENEPITOLEGPGPDLSGF 240

QY 241 TLNRQKAKHESLSKEIKRKGUHQHGPYTRTETTSMLIEHGHTPEAKSPBOLRDPG 300  
Db 241 TLNRQKAKHESLSKEIKRKGUHQHGPYTRTETTSMLIEHGHTPEAKSPBOLRDPG 300

RESULT 2

T24944  
hypothetical protein ZK1307.8 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
 C;Accession: T24944; T27735  
 R;Cardner, A.  
 submitted to the EMBL Data Library, January 1995  
 A;Reference number: Z19958  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: DNA  
 A;Residues: 1-507 <WIL>  
 A;Cross-references: EMBL:Z47356; PIDN:CAA87420.1; GSPDB:GN00020; CESP:ZK1307.8  
 A;Experimental source: clone T15H9  
 R;Barks, M.  
 submitted to the EMBL Data Library, January 1995  
 A;Reference number: Z20412  
 A;Accession: T27735  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: DNA  
 A;Residues: 1-507 <WIL>  
 A;Cross-references: EMBL:Z47358; PIDN:CAA87438.1; GSPDB:GN00020; CESP:ZK1307.8  
 A;Experimental source: clone ZK1307  
 C;Genetics:  
 A;Gene: CESP:ZK1307.8  
 A;Map position: 2  
 A;Lintrons: 34/1; 177/1; 265/1; 463/3  
 Query Match 11.4%; Score 185.5; DB 2; Length 507;  
 Best Local Similarity 38.7%; Pred. No. 6.7e-08; Mismatches 47; Indels 7; Gaps 4;  
 Matches 41; Conservative 11; Mismatches 47; Indels 7; Gaps 4;  
 QY 67 LSCKCFSIESTVYKECPFHNVTOHQFRMNAVSGI-LGIWHEME-IANTFTGMMR 124  
 DQ 394 LKQKCFDNRVQQTYQRPFGTQKOT---GAYGTSLSKPEKNSPKENKYKQHFG 449  
 QY 125 DGDAC-RSSRSQSKVELLAGKSNRLAHVSEPTSTCVALLFETPLVC 169  
 DQ 450 DGQCWNGWPKRSDTIDTEBGEENELVETPAKCELFTRPLAC 495  
 RESULT 3  
 J05889  
 OS-9 protein precursor - human  
 N;Contains: OS-9 protein, splice form 1; OS-9 protein, splice form 2; OS-9 protein, spliced  
 C;Species: Homo sapiens (man)  
 C;Date: 18-Mar-1998 #sequence\_revision 18-Mar-1998 #text\_change 22-Oct-1999  
 C;Accession: JCB889; JBR0106; JBR0107; JE0108  
 RKimura, Y.; Nakazawa, M.; Tsuchiya, N.; Asakawa, S.; Shimizu, N.; Yamada, M.  
 J; Biochem. 122, 1190-1195, 1977  
 A;Title: Genomic organization of the OS-9 gene amplified in human sarcomas.  
 A;Reference number: JCB889; MUID:98158329; PMID:9498564  
 A;Accession: JCB889  
 A;Molecule type: DNA  
 A;Residues: 1-667 <KIM1>  
 A;Cross-references: DDBJ:AB002806; NID:92780782; PIDN:BAA24363.1; PID:d1025275; PID:9278  
 A;Experimental source: sarcomas.  
 A;Note: neither the complete nucleic acid sequence nor the complete translation are shown  
 R;Kimura, Y.; Nakazawa, M.; Yamada, M.  
 J; Biochem. 123, 875-892, 1998  
 A;Title: Cloning and characterization of three isoforms of OS-9 cDNA and expression of t  
 A;Reference number: JEB0108; MUID:98230694; PMID:9562620  
 A;Accession: JE0106  
 A;Molecule type: mRNA  
 A;Residues: 1-667 <KIM2>  
 A;Cross-references: DDBJ:AB002806; NID:92780782; PIDN:BAA24363.1; PID:d1025275; PID:9278  
 A;Accession: JE0107  
 A;Molecule type: mRNA  
 A;Residues: 1-534, 590-667 <KIM3>  
 A;Cross-references: DDBJ:AB002806  
 A;Accession: JE0108  
 A;Molecule type: mRNA  
 A;Residues: 1-455, 471-534 590-667 <KIM4>  
 A;Cross-references: DDBJ:AB002806  
 C;Comment: This protein is involved in amplification and overexpression of various tumor

A;Gene: OS-9  
 A;Cross-references: GDB:998646  
 A;MPD position: 12G13.12Q15  
 C;Keywords: alternative splicing; carcinogenesis; glycoprotein  
 F;1-25/Domain: signal sequence #status predicted <SIG>  
 F;26-67/Domain: OS-9 protein, splice form 1 #status predicted <MAT1>  
 F;26-534,590-667/Product: OS-9 protein, splice form 2 #status predicted <MAT2>  
 F;26-455,471-534,590-667/Product: OS-9 protein, splice form 3 #status predicted <MAT3>  
 F;177/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 Query Match 9.5%; Score 155.5; DB 2; Length 667;  
 Best Local Similarity 21.7%; Pred. No. 3.6e-05;  
 Matches 79; Conservative 47; Mismatches 121; Indels 117; Gaps 18;  
 Db 21 PAGAKMKVTEEPNAFGVNPNFLPQASRIQAKEDPSPVSGPVHLFRUSKGCPFLVYSTYK 80  
 Db 76 PAGAIIHPQREREEBTAVQGPGLPFL--LSPMRD-----APCLIKTGDWT 119  
 QY 81 YEFCPFHNTQ-----HEQFRMWAVSGIGIWIWEIWEIANNTFGMM 123  
 Db 120 YEFCKGRHIQYEMDSETKGEVLYGGYQSAFDWDDETAKSKQRKLRKRYHSQTG-- 176  
 Db 177 -NGSKCDLNGRPREAEVFLCDEGAGISGSDYDIDRVDEPLCSYVLTTRPLCPHFLRP 235  
 QY 170 ---HPHALVYPTL-PFA---LQRQHDQE---ODLADEL-----ITPO- 203  
 Db 235 PPSAAPQATLCHSLSQPBEYMAVQRDASKQGDKTIEEQLDQHQWSETKSGAPQ 295  
 QY 204 -----GHEKLRLTFEDA-GYLKIP-TENEPT----QLEGADS-LGFETL 242  
 Db 296 MAGASPTKDDSKOSDFWQMLNPEDQAPGGERVPABQDPSPPEAASASGA-PNDQNNVQ 355  
 QY 243 ENCRKAHKBLSKIKRKLGKLTQHGIYTRPRETSNLHEHICHTP-----RAKSPEQL 295  
 Db 356 VKVIRSPADLIRFIELKG-GTKKGKD-----NIGOBQPDAAEVQPQREPEKE 403  
 QY 295 RGDP 299  
 Db 404 RGDP 407  
 RESULT 4  
 A32469 4  
 80K protein H precursor - human  
 C;Species: Homo sapiens (man)  
 C;Date: 12-Oct-1989 #sequence\_revision 12-Oct-1989 #text\_change 21-Jul-2000  
 C;Accession: A32469  
 A;Reference number: A32469; MUID:90007553; PMID:2793184  
 A;Status: preliminary  
 A;Molecule type: mRNA  
 A;Genomics: 5, 309-315, 1989  
 A;Title: Isolation of cDNAs encoding a substrate for protein kinase C: nucleotide sequence  
 A;Reference number: A32469; MUID:90007553; PMID:2793184  
 A;Accession: A32469  
 R;Sakai, K.; Hirai, M.; Minoshima, S.; Kudoh, J.; Fukuyama, R.; Shimizu, N.  
 A;Cross-references: GDB:PRKCSH; G19P1  
 A;Genes: GDB:PRKCSH; G19P1  
 A;Cross-references: GDB:PRKCSH; G19P1  
 A;Map position: 19p13.2-p13.1  
 C;Keywords: phosphoprotein  
 F;1-14/Domain: signal sequence  
 F;524-527/Region: endoplasmic reticulum retention signal  
 Query Match 8.6%; Score 140.5; DB 2; Length 527;  
 Best Local Similarity 30.5%; Pred. No. 0.0051;  
 Matches 32; Conservative 14; Mismatches 54; Indels 5; Gaps 3;  
 QY 67 LSCKCFSIESTVYKECPFHNVTOHQFRMNAVSGIGIWIWEIWEIANNTFGMMR 125  
 Db 410 LYSQCIVLTINNVYRLCPFLKLVQSQP--KLGGSPTLSGIGWGSWIGPDHDKFSAWKYEQ 466

QY 126 GDACTRSRSQVELAGKSNRLAHVSESTCVYALTFETPLVC 169  
Db 467 GTGCWQSPNRSITIVLRLGKEMVISTEPSRCVILMELMPAAC 511

**RESULT 5**

T01158 hypothetical protein F7N22.4 - Arabidopsis thaliana (mouse ear cress)  
C;Species: Arabidopsis thaliana (mouse ear cress)  
C;Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 21-Jan-2000  
C;Accession: T01158  
R;Dante, M.  
submitted to the EMBL Data Library, April 1998  
A;Description: The sequence of A. thaliana F7N22.  
A;Reference number: Z14250  
A;Accession: T01158  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-150 <DNA>  
A;Cross-references: EMBL:AF058825; NID:93047060; PID:93047065  
C;Genetics:  
A;Experimental source: cultivar Columbia  
A;Map position: 4  
A;Introns: 62/3; 92/2;  
A;Note: FN22.4  
A;Superfamily: Arabidopsis thaliana hypothetical protein F7N22.4

Query Match 6 5%; Score 106; DB 2; Length 150;  
Best Local Similarity 25.4%; Pred. No. 0.084; Mismatches 49; Indels 36; Gaps 7;  
Matches 35; Conservative 18;

QY 45 QASRLOQAKRDSPVSPGVPHLFRISKGCPFLSPLVSTYKVEFC-----PFHNTQHEQTFRW 98  
Db 38 STQOLVKLKTPEDELQG----LSBKCLFRQEGWNSYEFHQKVYRQLHEDENKVGFYN 92

QY 99 NAYSGIGIWHWEIANNT--FTGKMRDGACRSRSQVELAGKSNRL-AHVSEF 154  
Db 93 -----YISHVYINGTCDLTG-----SPRVEVRVCAETRAMTISITE 132

QY 155 SNCVYALTFETLVC-HP 171  
Db 133 STCKYALTVQCPTICKHP 150

**RESULT 6**

F70318 hypothetical protein aq\_199 - Aquifex aeolicus  
C;Species: Aquifex aeolicus  
C;Accession: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 17-Mar-2000  
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov  
v. Nature 392, 353-358, 1998  
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A;Reference number: A70300; MUID:98:9666; PMID:957320  
A;Accession: F70318  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-407 <RDP>  
A;Cross-references: GB:AE000678; NID:92982921; PID:AC06549.1; PID:92982932; GB:AB00065  
C;Genetics:  
A;Gene: aq\_199  
C;Superfamily: translation releasing factor eRF-1

Query Match 5.8%; Score 95; DB 2; Length 407;  
Best Local Similarity 20.6%; Pred. No. 2.8;  
Matches 58; Conservative 33; Mismatches 95; Indels 96; Gaps 14;

QY 42 FLPQASRLOQAKRD-PSPVSPGVPHLFRISKGCPFLSPLVSTYKVEF-----CHPHNTQHEQT 95  
Db 147 FLMDITGIEBKUDLFPMTRAHFRHSGGALLKGAGCTFQYRMPARGAAP-NVVQHGNG 204

QY 96 FRNAYASGILGIGIWE-WEIAANNFTGWMNRDGDACRSRSQVELAGKSNRLAHVSEF 154  
Db 205 -EWRPHKIREWHRWIKLADSAFEEWKR----- 233

QY 155 STCWYALTFETPLVCHHALTYPTPEALORQWQDQEVLQDLADELITPQGHKEKLRTLFE 214  
Db 234 -----IHDKLIGGFBEGLE-----IENFLIPYVKEKLV----- 265

QY 215 DACYLK-TPENEPTO-EGGDSLGFPTELNCRAKHKELSKEIKLGL-LTOG----- 267  
Db 266 --GYIETTPPEAMPHQMEKADLL---WQDRDEQEKEIKEELBELKGWGLAVNGTSE 319

QY 268 -----IP-YTRPETSNSHLGHBTPRAKSPQ 294  
Db 330 LEMLAIGNVRTLIPPERFKCGYLQOSHSLAFLKPCPCLIE 361

**RESULT 7**

A48343 24K hypothetical protein - chicken anemia virus (isolate Cux-1)

C;Species: chicken anemia virus, CAV  
C;Accession: A48343  
R;Meehan, B.M.; Todd, D.; Creelan, J.L.; Earle, J.A.; Hoey, B.M.; McNulty, M.S.  
Arch. Virol. 124, 301-319, 1992  
A;Title: Characterization of viral DNAs from cells infected with chicken anaemia agent:  
A;Reference number: A48343; MUID:9239689; PMID:1605740  
A;Accession: A48343  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-216 <MEB>  
A;Cross-references: GB:MB1223; NID:9323254; PID:AAA42882.1; PID:9323255  
A;Note: sequence extracted from NCBI backbone (NCBIN:106168, NCBIPI:106169)

Query Match 5.7%; Score 93.5; DB 2; Length 216;  
Best Local Similarity 19.6%; Pred. No. 1.6; Mismatches 54; Conservative 43; Indels 103; Gaps 14;  
Matches 54; Conservative 43; Mismatches 75; Indels 103; Gaps 14;

QY 17 GGPPAGAKMKVVEERPAFGVNNPFHQASLQLAKERDPSPVSPGPPVHLFRISGKCFSLVE 76  
Db 21 GQGPSSGAQGVISN----- 216  
Db 21 GQGPSSGAQGVISN----- 216  
Db 77 STYKVECPFHNTQHEOTFRNAYASGILGIGIWE-WEIAANNFTGWMNRDGDAC 129  
Db 53 ATNKFTAVGNPSIQORDPDMYRM-YMISIAWNRGESSRSHAKICNGQQFRKWFQECAGL 111

QY 130 RSRSRQVELAGKSNRLAHVSESTCVYALTFETPLVCHHALTYPT 179  
Db 112 EDRSTQASLLEAIRLRPLRVQGRKRAKRLDKYHSQPT----- 156  
C;Accession: F70318 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov  
v. Nature 392, 353-358, 1998  
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A;Reference number: A70300; MUID:98:9666; PMID:957320  
A;Accession: F70318  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-407 <RDP>  
A;Cross-references: GB:AE000678; NID:92982921; PID:AC06549.1; PID:92982932; GB:AB00065  
C;Genetics:  
A;Gene: aq\_199  
C;Superfamily: translation releasing factor eRF-1

Query Match 5.8%; Score 95; DB 2; Length 407;  
Best Local Similarity 20.6%; Pred. No. 2.8;  
Matches 58; Conservative 33; Mismatches 95; Indels 96; Gaps 14;

QY 42 FLPQASRLOQAKRD-PSPVSPGVPHLFRISKGCPFLSPLVSTYKVEF-----CHPHNTQHEQT 95  
Db 147 FLMDITGIEBKUDLFPMTRAHFRHSGGALLKGAGCTFQYRMPARGAAP-NVVQHGNG 204

QY 96 FRNAYASGILGIGIWE-WEIAANNFTGWMNRDGDACRSRSQVELAGKSNRLAHVSEF 154  
Db 205 -EWRPHKIREWHRWIKLADSAFEEWKR----- 233

QY 155 STCWYALTFETPLVCHHALTYPTPEALORQWQDQEVLQDLADELITPQGHKEKLRTLFE 214  
Db 234 -----IHDKLIGGFBEGLE-----IENFLIPYVKEKLV----- 265

QY 215 DACYLK-TPENEPTO-EGGDSLGFPTELNCRAKHKELSKEIKLGL-LTOG----- 267  
Db 266 --GYIETTPPEAMPHQMEKADLL---WQDRDEQEKEIKEELBELKGWGLAVNGTSE 319

QY 268 -----IP-YTRPETSNSHLGHBTPRAKSPQ 294  
Db 330 LEMLAIGNVRTLIPPERFKCGYLQOSHSLAFLKPCPCLIE 361

**RESULT 8**

P86904 conserved hypothetical protein YXCA [imported] - Lactococcus lactis subsp. lactis (stra  
C;Species: Lactococcus lactis subsp. lactis  
C;Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
C;Accession: P86904  
R;Bolotin, A.; Wincker, P.; Mager, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrl  
Genome Res. 11, 731-733, 2001  
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s  
A;Reference number: A86625; MUID:21235185; PMID:11337471  
A;Accession: F86904  
A;Status: preliminary  
A;Molecule type: DNA

A;Residues: 1-1435 <S90>  
A;Cross-references: GB:AE005176; PID:912725306; PIDN:AAK06336.1; GSPDB:GN00146  
A;Experimental source: strain IM1403  
A;Gene: yxca

Query Match 5.7%; Score 93.5; DB 2; length 1435;  
Best Local Similarity 27.3%; Pred. No. 21; Mismatches 47; Conservative 21; MisMatches 55; Indels 49; Gaps 9; Matches 34; Conservative 25; Mismatches 40; Indels 38; Gaps 4; Db 959 NKKYTVLKLIDEGSGNGAATRIRSIKAV--SERTRNTEASTZPHELVQETPFTKE 1014  
Qy 168 VCHPHALIWPPTPEALQRQWDOVEQDILADELTPOGHKURLTLEDAGY--LKTPBEN 225  
Db 1015 MAKGTTL--LP-----MLSPHQEGILDATPAAAGGNVSLPESN 1053  
Qy 226 EPFLQLEGDPDSLGPFETENCRKAHKELSKETIKRLKG--LLTIG 267  
Db 1054 --TSVNNG---LKVNNDNSCYPAIITIGQLIBALQSGEYDIDNTSVMMTQG 1100

RESULT 9

T3152 hypothetical protein - fission yeast (*Schizosaccharomyces pombe*) (fragment)

C;Species: *Schizosaccharomyces pombe*  
CDat: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 21-Jul-2000  
CAccesion: T43152  
R;Yoshitaka, S.; Kato, K.; Nakai, K.; Okayama, H.; Nojima, H.  
DNA Res 4, 363-369, 1997

A;Title: Identification of open reading frames in *Schizosaccharomyces pombe* cDNAs.  
A;Reference number: Z17323; MUID:98162722; PMID:9501991  
A;Accession: T43152

A;Status: preliminary; translated from GB/EMBL/DDJB  
A;Residues: 1-151 <Y05>  
A;Cross-references: EMBL:DB9245; NID:91749597; PIDN:BAA13906.1; PID:91749598  
A;Experimental source: strain PR745

Query Match 5.6%; Score 92; DB 2; Length 515;  
Best Local Similarity 21.9%; Pred. No. 7; Mismatches 25; Conservative 16; MisMatches 47; Indels 26; Gaps 3; Matches 26; Db 87 HNTQHEQTPR--WNAVSGLGIWHWEIANNFTGMW----- 122

Qy 373 HTKLKHLBELEYHGWDLTAKIOMKETREKIGGYTKWVYENFQPSILGNFASQEGNTL 432  
Db 123 -MEDGDAQ-RSRSRQSRSKVELAGKSNRHLAHVSEPSCTVYALTFTPLVCHHAL 174  
Qy 433 KYENGQSCWNGPHRSAITVTCGVENBIVSLEAQCKCEYIYIKMSPAACSPNQL 486  
Db 433 KYENGQSCWNGPHRSAITVTCGVENBIVSLEAQCKCEYIYIKMSPAACSPNQL 486

RESULT 10

I36911 involucrin L - douroucouli  
C;Species: *Atotus trivirgatus* (douroucouli, night monkey, owl monkey)  
CDat: 16-Feb-1996 #sequence\_revision 23-Feb-1996 #text\_change 22-Jun-1999  
CAccesion: I36911  
R;Tseng, H.; Green, H.  
Mol. Biol. Evol. 6, 460-468, 1989

A;Title: The involucrin gene of the owl monkey: origin of the early region.  
A;Reference number: I36911; MUID:90014142; PMID:2507864  
A;Status: nucleic acid sequence not shown  
A;Molecule type: DNA  
A;Residues: 1-1544 <RSS>

A;Cross-references: GB:M25313; NID:9176557; PIDN:AAA35375.1; PID:9176558  
C;Comment: During the terminal differentiation of keratinocytes, this protein from the c-linked envelope under the plasma membrane.

C;Superfamily: involucrin

C;Keywords: cornified cell envelope; duplication; epidermis; tandem repeat

C;Accession: F153-501/Region: 10-residue repeats (Q-E-G-Q-[PLV]-[KE]-[LH]-[PL]-E-Q)

Query Match 5.6%; Score 92; DB 1; Length 544;  
Best Local Similarity 23.4%; Pred. No. 7; Mismatches 34; Conservative 25; Mismatches 40; Indels 38; Gaps 4; Db 110 EKAQREKQOLQGQEBEKULLDQDPMHAKLSDQEQGKLEQQ 169  
Qy 182 EALORQWDQVEQDIA-----LITOGHEKLRUTIFED 215  
Db 216 AGYKLTPENEPTOLEGPDPDSLGPETLENCRKAHKELSKETIKRICKLITPHGIPYTRPT 274  
Db 170 EGHLIELPEQEGQ-----LKCLBQ-OEQHQLPEQOBGQLKHQBEGQQLKHLQ 218  
Qy 275 ETSNLHUGHETRAKSPEQLQDP 299  
Db 219 QEGQVKHLQEQKOSELPEQOQGP 243

RESULT 11

C59436 KIAA1391 protein [imported] - human

C;Species: Homo sapiens (man)  
CDat: 03-Jun-2002 #sequence\_revision 03-Jun-2002 #text\_change 23-Sep-2002

R;Nagase, T.; Kikuno, R.; Ishikawa, K.I.; Hirose, M.; Ohara, O.

DNA Res. 7, 65-73, 2000  
DNA Res. 7, 65-73, 2000  
A;Title: Prediction of the coding sequences of unidentified human genes. XVI. The compl

A;Reference number: D59436  
A;Accession: C59436

A;status: preliminary

A;Molecule type: mRNA

A;Residues: 1-1194 <NAG>

A;Cross-references: GS:BA92629; PID:9743163; PIDN:BAA92629.1

R;Ohara, O.; Nagase, T.; Kikuno, R.

submitted to Genbank, January 2000

A;Reference number: D59436

A;Accession: D59436

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-1194 <OHA>

A;Cross-references: GB:BAA92629; PID:9743163; PIDN:BAA92629.1

Query Match 5.6%; Score 92; DB 2; Length 1194;  
Best Local Similarity 22.2%; Pred. No. 22; Mismatches 63; Conservative 40; MisMatches 103; Indels 78; Gaps 13; Matches 63; Db 64 LFRL---SGKCSLIVESTYKFP---CPFHNTQHETQFRWNAISGILGWHWEIANN 116

Qy 14 LSAGGAPAGA----AKMKVVEPAFMAGVNNPFHQASRQLAKR--DPSP-VSGPVH 63  
Db 941 LSSPGTSPSGSSVSSQDSAFSOISEHSVFTPTTSPIDCTFOAQOKREDLSPDFSNASH 1000

Qy 64 LFRL---SGKCSLIVESTYKFP---CPFHNTQHETQFRWNAISGILGWHWEIANN 116

Db 1001 VSMPGPSSQGACSRPRTKDTMENHSQMHSVTPST----- 1039

Qy 117 TFTGMWQRDGDA-CRSRSRQSRSKVELAGKSNRHLAHVSEPSCTVYALTFTPLVCHHAL 175  
Db 104 ---WLRNGVSLKWSKNSKAKARPEEKIASKG-----LEP---PHA-- 1080

Qy 176 VYPTPEALQRQWDOVEQD-----ADELTPOGHKURLTLEDAGY--LKTPBEN 233  
Db 1081 -SGVPEANSLOBEQDPLRRAEGLSPVQAOCRQCSSPFDQ-----ERHCSS 1127

Qy 234 PSLGFFELNENCRKAHKELSKETIKRICKLITPHGIPYTRPT 277  
Db 1128 PSLVLESRKLCMKSKHIEPEGSQSSG-----SLPRERASASS 1166

Query Match 5.6%; Score 92; DB 1; Length 544;  
Best Local Similarity 23.4%; Pred. No. 7; Mismatches 34; Conservative 25; Mismatches 40; Indels 38; Gaps 4; Db 110 EKAQREKQOLQGQEBEKULLDQDPMHAKLSDQEQGKLEQQ 169  
Qy 182 EALORQWDQVEQDIA-----LITOGHEKLRUTIFED 215  
Db 216 AGYKLTPENEPTOLEGPDPDSLGPETLENCRKAHKELSKETIKRICKLITPHGIPYTRPT 274  
Db 170 EGHLIELPEQEGQ-----LKCLBQ-OEQHQLPEQOBGQLKHQBEGQQLKHLQ 218  
Qy 275 ETSNLHUGHETRAKSPEQLQDP 299  
Db 219 QEGQVKHLQEQKOSELPEQOQGP 243

RESULT 12

A39926 hypothetical protein 1 - chicken anemia virus

C;Species: chicken anemia virus, CAV  
CDat: 17-Jan-1992 #sequence\_revision 17-Jan-1992 #text\_change 08-Oct-1999

C;Accession: A39926

R. Noteborn, M.H.M.; de Boer, G.F.; van Roozelar, D.J.; Karreman, C.; Kranenburg, O.; V.J. Virol. 65, 3131-3139, 1991  
A.Title: Characterization of cloned chicken anemia virus DNA that contains all elements  
A.Reference number: A39926; MUID:91257831; PMID:1811873  
A.Accession: A39926  
A.Status: preliminary  
A.Molecule type: DNA  
A.Residues: 1-216 <NOT>  
A.Cross-references: GB:M5918; NID:9323250; PIDN:AAA91822.1; PID:9323251  
C:Comment: This virus is unclassified.

RESULT 14  
T30545 major surface glycoprotein - *Pneumocystis carinii* (fragment)  
C.Species: *Pneumocystis carinii*  
C.Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 15-Jun-2001  
C;Accession: T30545  
R;Mei, Q.; Turner, R.E.; Serial, V.; Klivington, D.; Angus, C.W.; Kovacs, J.A.  
Infect. Immun. 66: 4288-4273, 1998  
A;Title: Characterization of major surface glycoprotein genes of human *Pneumocystis carinii*.  
A;Reference number: Z17905; MUID:98880374; PMID:9712777

Query Match Similarity 5.6%; Score 91.5; DB 2; Length 216;  
 Best Local Similarity 19.3%; Pred. No. 2.4; Matches 53; Conserv. 44; Mismatches 75; Indels 103; Gaps 14;

A-Accession: T3045  
A-Status: preliminary; translated from GB/EMBL/DDJB  
A-Molecule type: DNA  
A-Segments: 1-1014 <MET>  
A-Cross-references: EMBL:AB038556; NID:93560524; PMID:93560525; PIDN: AAC34980.1

C;Experimental source: f.sp. hominis  
 C;Genetics:  
 C;Superfamily: *Pneumocystis carinii* major surface glycoprotein MSG100  
 Query Match 5.6%; Score 91.5; DB 2; Length 1014;  
 Best Local Similarity 19.6%; Pred. No. 19;  
 Matches 58; Conservative 42; Mismatches 13; Indels 65; Gaps 10;

QY	49	LOAKRDPSPVSGPVHLFRISKGKFLSVSTYKEFCPHNVTIH-EQTFRWNAISGILGI	107
Db	524	LNRKD-FPKEDCE-ELIKCGEALRDKSKKEW-PCHTLSONCDORNAKELKELLIN	579

157 V-----RW-----QDELDAD-----RBADF-----TPSEE-----DGTTSSDP 184  
 240 ETLENCRKAHKELSKERKRLKGULIO-HGIPYTRP 273  
 185 D-----RDNFTGGSGTGTGPTMGPRT 210

QY 108 WHEWELIANNTFTGMWMRDODACRSRSROSKVELACKSNSRLAHVSEPSITCVALTETPL 167  
 :  
 Db 580 EHKOILKHOENCNGMYLKEKCNENSRRRNERFESLCAQNTRGRIMEDVKNOCKIFENI 639  
 QY 168 - - - - - VCHPHALLVYPTLPEALQRQWDOVE - - - - - QDLADE 198

RESULT 13  
T41623  
probable protein kinase C substrate - fission yeast (*Schizosaccharomyces pombe*)

;Species= Schizosaccharomyces pombe  
;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
;Accession: TU1629  
;Accession: TU1629  
;Aert, R.; Volckaert, G.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL, DDBJ, GenBank, October 1999

QY	242	L ENR----- KAHELSKEIKRUGGLTOH:G BYTRPIETS N FHGL 283
Db	760	KERQGGLSKELEAKNDLEEKHDYKTYKTDKNAAMBNTLYFSTKSTDNKTEKG 815

Source: Reference number: 221735  
Accession: 141623  
Status: preliminary; translated from GB/EMBL/DDJB  
Molecule type: DNA  
Primer: 5'-ATG-3'  
->5'-TCA-3'

RESULT 15  
T47587  
TATA box binding protein (TBP) associated factor (TAF)-like protein - Arabidopsis thaliana  
Name: protein\_T2A8P22\_240

;Map Position: 3  
;Strain: 972h-  
;Cross-references: EMBL:AL122011; PIDN: CAB58410\_1; GSPDB: GN00068; SPDB: SPCCB25\_02  
;Experimental source: strain 972h-; cosmid c825  
;Genetics:  
;Gene: SPDB:SPCCB25\_02

RESULT 15  
T7587 TATA box binding protein (TBP)-like protein - Arabidopsis thaliana  
N;Littermate names: protein\_F24B2.240  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
C;Accession: T47587  
R;Bloemberg, H.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.; Quertier, F.; Salanoubat, M.  
submitted to the Protein Sequence Database, January 2000

Quarry March 5% Score 91 DR 2 Length 2049.  
A;Accession: T47587  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-2049 <BIO>  
A;Cross-references: EMBL:All32957  
A;Experimental source: cultivar Columbia; BAC clone F24B22  
C;Genetics:  
A;Map Position: 3  
A;Introns: 18/1; 49/3; 122/2; 140/3; 168/1; 280/3; 314/1; 512/2; 556/3; 579/3; 550/2; 7  
A;Note: /3; 1995/3  
A;Note: F24B22.240

Db 976 APPTKDIELPVAELSRITVIM-RNEASQQLHTE-TCHCFDYLSTINLN--VESVAD 1030  
QY 93 EOPFRWNAYSGIGIWHWEIANNTFGQNMWRDGDACTRSRSRQSKYELACCSNRLAHVS 152  
Db 1031 ETI---DEASTLDIWNESAGNESLKVQVFEDVESSRQULISTAGYLCVQSN--IHT 1084  
QY 153 EPSTCWLALTF--ETPLVCHPHALVYPTLPALQRQWQVQDPLADELT----- 201  
Db 1085 VTSVLVAAAVWWMSBPPARLNP---ITLP-LMASIKRQEBOILQIAAEALELIAYCVR 1140  
QY 202 -POCHEKLRT-----LFEDAGYL---KTPPEENEPTELEGGP 234  
Db 1141 KRPFDNLKJKNICSLTCMDPSETPOASIISSMDIVDDMFSSRSITGKOKAVVYLASGE 1200  
QY 235 DSL---GBTLENCRKAKELS-----KEIKRKLGLTONGIP 269  
Db 1201 DRSKVEGFTTRSELAKHLISLKFGSSLFDKLPKLMCECLTELVLP 1246

Search completed: July 26, 2004, 11:12:26  
Job time : 9.2294 secs



CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR U41635; AA06495.1; --.  
 DR EMBL; AB002805; BRA24562.1; --.  
 DR EMBL; AB002806; BRA24563.1; --.  
 DR EMBL; UB1031; AAC39523.2; --.  
 DR PIR; JCS589; JCS589.  
 DR GO; GO:0008452; C:soluble fraction; TAS.  
 DR GO; GO:0008451; D:cell growth and/or maintenance; TAS.  
 DR InterPro; IPR009011; Man\_6\_P\_R bind.  
 KW SIGNAL; Alternative splicing; Polymorphism.  
 FT VARSPLIC 1 25  
 FT CHAIN 26 667  
 FT DOMAIN 414 429  
 FT CARBOHYD 177 177  
 FT VARSPLIC 456 470  
 FT VARSPLIC 535 589  
 FT VARIANT 398 398  
 FT SEQUENCE 667 AA; 75561 MW; 65BA3F66C5C58756 CRC64;  
 QY 21 PAGAKMKVVEEINAFGNNPFLPQASRLQATRDPSPVGVPVHLRLSGKCFSLIVESTYK 80  
 DB 76 PAGAIIHFERERETPAYQGPPIPEL-LSPMRD-----APCLLKTKDWMT 119  
 DB 81 YECPPFHNTQ-----HEQTRWAVAYSGIGLIGIWHEWELIANNFTGMM 123  
 DB 120 YFCYGRHHIQQYHEDMSBKEIGEVLYNGYYQSADFWMDDETAKSKOIRLKRYHSOTYG-- 176  
 QY 124 RDGJAC-RSRSRQSKVELACK-----SNKLAHVAEPSTVWALTFETAVC----- 169  
 DB 177 -NSKSCDINGRPREAEVFRFLCPFGAGGSSGYIDRVDEPLSCSVWLTTRPLCPHPLRP 235  
 QY 170 ---HPHALLVYPTL-BPA---LQRQDQVE-----QDADELL----ITPQ- 203  
 DB 236 PRSARAPQILCHISLQFREMYAVQYRQDSKGDKLLEELQDGQWMSRTKSYAVPK 295  
 QY 204 -----GHEKLRLTFEDA-GYLKTP-SENKEPT----OLESGPDSLGFTL 242  
 DB 296 MAGASPTKDDSKSDFWMLNPFEDQPGGERVPAEQQDPSPEDAASGAPNDFQNNVQ 355  
 QY 243 ENRKAKHESLKEIKRIGLILQHGIYTRPETSNEHIGHETP-----RAISPEQL 295  
 DB 355 VKVIRSPADLIRFIEELKG-GTKKGKP-----NIGOBQPVDAAEVPOREPEKE 403  
 QY 26 RGDP 299  
 DB 404 RGDP 407  
 RESULT 2  
 G19P\_BOVIN STANDARD; PRT; 533 AA.  
 ID\_G19P\_BOVIN STANDARD; PRT; 527 AA.  
 AC P14314;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JUL-1999 (Rel. 38, Last annotation update)  
 DE Protein kinase C substrate, 80 kDa protein, heavy chain (PKC<sub>SH</sub>)  
 DE (80K-H Protein).  
 GN PRKCSH.  
 OC Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=913;  
 RN [1];  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE:90007553; Published=279314;  
 RA Sakai K., Masamichi H., Minoshima S., Kudoh J., Fukuyama R.,  
 RA Shimizu N., "Isolation of cDNAs encoding a substrate for protein kinase C:  
 RT nucleotide sequence and chromosomal mapping of the gene for a human  
 RT 80K protein.",

RL Genomics 5:309-315(1989).  
 RN [2]  
 SQUENCE FROM N.A.  
 RX MEDLINE=97196775; PubMed=9043864;  
 RA Ophoff R.A., Terwindt G.M., Vergouwe M.N., van Eijk R., Ferrari M.D.,  
 RA Mohrweiser H., Litt M., Hofker M.H., Haan J., Ferrari M.D.,  
 RA Prants R.R.;  
 RT "A 3-Mb region for the familial hemiplegic migraine locus on 19p13.1-  
 P13.2; exclusion of PRKCSH as a candidate gene.";  
 RL Eur. J. Hum. Genet. 4:321-328(1996).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement.(See <http://www.isb-sib.ch/announce/>  
 or send an email to license@isb-sib.ch).  
 CC  
 CC  
 DR EMBL; J03075; AAA52493.1; --.  
 DR EMBL; U50326; AAA98688.1; --.  
 DR EMBL; U50317; AAA98688.1; JOINED.  
 DR EMBL; U50318; AAA98688.1; JOINED.  
 DR EMBL; U50319; AAA98688.1; JOINED.  
 DR EMBL; U50320; AAA98688.1; JOINED.  
 DR EMBL; U50321; AAA98688.1; JOINED.  
 DR EMBL; U50322; AAA98688.1; JOINED.  
 DR EMBL; U50323; AAA98688.1; JOINED.  
 DR EMBL; U50324; AAA98688.1; JOINED.  
 DR EMBL; U50325; AAA98688.1; JOINED.  
 DR PIR; A32469; A32469.  
 DR MM; 177060; .  
 DR GO; GO:0005622; C:intracellular; NAS.  
 DR Interpro; IPR00243; EF-hand.  
 DR Interpro; IPR00886; ER target.S.  
 DR Interpro; IPR002172; LDL receptor.A.  
 DR Pfam; PF00356; Efhand\_2.  
 DR SMART; SM0192; LDDA; 1.  
 DR PROSITE; PS00014; ER TARGET; 1.  
 DR PROSITE; PS00018; EF\_HAND; 1.  
 KW Phosphorylation.  
 FT DOMAIN 313 336 GLU-RICH (ACIDIC)  
 MOD\_RES 89 89 PHOSPHORYLATION (BY PKC) (POTENTIAL).  
 MOD\_RES 382 382 PHOSPHORYLATION (BY PKC) (POTENTIAL).  
 MOD\_RES 389 389 PHOSPHORYLATION (BY PKC) (POTENTIAL).  
 SQ SEQUENCE 527 AA; 59296 MW; 2BB2F9AC14E3FC94 CRC64;  
 Query Match. 8-6\*: Score 140.5; DB 1; Length 527;  
 Best Local Similarity 30.5%; Pred. No. 0.00033; Length 527;  
 Matches 32; Conservative 14; Mismatches 54; Indels 5; Gaps 3;  
 Qy 67 LSGKCFPLVEESTYKVERCPFHNTQHQETFRNNAVSGILJWHEN-BIANNTFTGWWRD 125  
 Db 410 IYSQCYLITNTVYVYLCPFLVSKP--KLGGSTTSIIGWGSPIGDIDKFAMKVQ 466  
 Qy 126 GDA-C-RERSRSRDKVVELAGKSRLAHVSEPSCTCVALTETPLVC 169  
 Db 467 GTCGWCQSPRSRSTVRLCKGKEMWTSITE-SRCBVLMEMLWPAAC 511  
 RESULT 4  
 K2L4\_HUMAN ID\_K2L4\_HUMAN STANDARD; PRT; 377 AA.  
 AC 099706; O14621; 014622; 014623; O14624; 043534; P78400; P78401;  
 AC 099559; 099560; 099561; 099562; Q9QQT7;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2003 (Rel. 42, Last annotation update)  
 DE Killer cell immunoglobulin-like receptor 2D4 precursor (MHC class I  
 DE NK cell receptor KIR103AS) Killer cell inhibitory receptor 103AS)  
 DE (KIR-103AS) (G9P);  
 GN KIR2D4 OR KIR103AS.  
 OS Homo sapiens (Human).  
 OC Bucaryota; Metacoda; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SQUENCE FROM N.A.  
 RX MEDLINE=97102173; PubMed=8946682;  
 RA Selvakumar A., Steffens U., Dupont B.;  
 RT "NK cell receptor gene of the KIR family with two Ig domains but  
 highest homology to KIR receptors with three Ig domains.";  
 RT Tissue Antigens 48:285-294(1996).  
 RN [2]  
 RP SQUENCE FROM N.A. (ISOFORMS 1; 3; 4 AND 5).  
 RX MEDLINE=97308880; PubMed=9234477;  
 RA Selvakumar A., Steffens U., Palanisamy N., Chaganti R.S.K., Dupont B.;  
 RT "Genomic organization and allelic polymorphism of the human killer  
 cell inhibitory receptor gene KIR103.;"  
 RL Tissue Antigens 49:564-573(1997).  
 RN [3]  
 RP SQUENCE FROM N.A. (ISOFORMS 1; 3 AND 6).  
 RC TISSUE=Lymphoid;  
 RA Bissonni R.;  
 RT Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SQUENCE FROM N.A. AND VARIANTS.  
 RX MEDLINE=980006; PubMed=9430221;  
 RA Unberg M., Valiente N.M., Shum B.P., Shilling H.G.,  
 RA Liebert Weidenbach K., Corliss B., Tyian D., Lanier L.L., Parham P.;  
 RT "Human diversity in killer cell inhibitory receptor genes.";  
 RL Immunity 7:753-763(1997).  
 RN [5]  
 RP SQUENCE FROM N.A.  
 RA Chan H.W., Salter R.D.;  
 RT "Exon deletion contributes to structural diversity of 2D4 killer  
 receptor isoforms.";  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SQUENCE FROM N.A. (ISOFORM 2).  
 RX MEDLINE=97213129; PubMed=059894;  
 RA Selvakumar A.;  
 RT "Polymorphism and domain variability of human killer cell inhibitory  
 receptors.";  
 RL Immunol. Rev. 155:183-196(1997).  
 CC INHIBITS THE ACTIVITY OF NK CELLS THUS PREVENTING CELL LYSIS.  
 CC -I- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -I- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=6;  
 CC Name=1; IsoId=099706-1; Sequence=Displayed;  
 CC Name=2; Synonyms=AST;  
 CC IsoId=099706-2; Sequence=VSP\_002609;  
 CC Name=3; Synonyms=AS;  
 CC IsoId=099706-3; Sequence=VSP\_002610;  
 CC Name=4; Synonyms=ASD1;  
 CC IsoId=099706-4; Sequence=VSP\_002609, VSP\_002610;  
 CC Name=5; Synonyms=ASD2;  
 CC IsoId=099706-5; Sequence=VSP\_002609, VSP\_002610, VSP\_002611;  
 CC Name=6; IsoId=099706-6; Sequence=VSP\_002608, VSP\_002609, VSP\_002610;  
 CC -I- SIMILARITY: Belongs to the immunoglobulin superfamily.  
 CC -I- SIMILARITY: Contains 2 immunoglobulin-like C2-type domains.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement.(See <http://www.isb-sib.ch/announce/>  
 or send an email to license@isb-sib.ch).

CC  
DR EMBL; U71199; AAC49755.1; -.  
DR EMBL; AF003123; AAB61926.1; -.  
DR EMBL; AF003116; AAB61926.1; JOINED.  
DR EMBL; AF003117; AAB61926.1; JOINED.  
DR EMBL; AF003118; AAB61926.1; JOINED.  
DR EMBL; AF003119; AAB61926.1; JOINED.  
DR EMBL; AF003121; AAB61926.1; JOINED.  
DR EMBL; AF003122; AAB61926.1; JOINED.  
DR EMBL; AF003120; AAB61926.1; JOINED.  
DR EMBL; AF002979; AAB71387.1; -.  
DR EMBL; AF002980; AAB71388.1; -.  
DR EMBL; AF002981; AAB71389.1; -.  
DR EMBL; AF002982; AAB71390.1; -.  
DR EMBL; X97229; CAAG65862.1; -.  
DR EMBL; X99479; CAAG67842.1; -.  
DR EMBL; X99480; CAAG67843.1; -.  
DR EMBL; X99481; CAAG67844.1; -.  
DR EMBL; AF034771; AAB5164.1; -.  
DR EMBL; AF034772; AAB5165.1; -.  
DR EMBL; AF034773; AAD24763.1; -.  
DR EMBL; AF110035; AAD24764.1; JOINED.  
DR EMBL; AF110332; AAD24763.1; JOINED.  
DR EMBL; AF110034; AAD24763.1; JOINED.  
DR EMBL; U73394; AAC51166.1; -.  
DR Genew; HGNC:6332; KIR2DL4.  
DR MIM; 604945; -.  
DR GO; GO:0005887; C:integral to plasma membrane; TAS.  
DR GO; GO:0005888; P:cellular defense response; TAS.  
DR GO; GO:0007165; P:signal transduction; TAS.  
DR InterPro; IPR001110; Ig-like.  
DR PR00007; Ig.2.  
DR PROSITE; PS0835; Ig\_LIKE; FALSE\_NEG.  
KW Receptor; Immunoglobulin domain; Glycoprotein; Signal; Transmembrane;  
KW repeat; Multigene family; Alternative splicing; Polymorphism.  
FT SIGNAL 1 21 BY SIMILARITY.  
FT CHAIN 22 KILLER CELL IMMUNOGLOBULIN-LIKE RECEPTOR  
FT D1SULFID 51 2014.  
FT DOMAIN 22 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 243 POTENTIAL.  
FT DOMAIN 264 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 44 IG-LIKE C2-TYPE 1.  
FT DOMAIN 139 IG-LIKE C2-TYPE 2.  
FT D1SULFID 51 BY SIMILARITY.  
FT DISULFID 146 BY SIMILARITY.  
FT DISULFID 146 195 N-LINKED (GLCNAC. . ) (POTENTIAL).  
FT CARBOHYD 141 N-LINKED (GLCNAC. . ) (POTENTIAL).  
FT CARBOHYD 175 Missing (in isoform 6).  
FT VARSPLIC 27 /FTId:VSP\_002608/  
FT VARSPLIC 219 Missing (in isoform 2, isoform 4, isoform 5 and isoform 6).  
FT VARSPLIC 235 /FTId:VSP\_002609/  
FT VARSPLIC 236 Missing (In isoform 3, isoform 4, isoform 5 and isoform 6).  
FT VARSPLIC 271 Missing (In isoform 5).  
FT VARSPLIC 271 /FTId:VSP\_002610/  
FT VARIANT 53 Y -> C (In dbSNP:618835).  
FT VARIANT 87 L -> V /FTId:VAR\_010307.  
FT VARIANT 138 T -> A /FTId:VAR\_010308.  
FT VARIANT 138 /FTId:VAR\_010309.  
FT VARIANT 209 A -> P (In dbSNP:1051456).  
FT VARIANT 271 D -> N /FTId:VAR\_010310.  
FT VARIANT 371 H -> N /FTId:VAR\_010311.  
FT VARIANT 371 /FTId:VAR\_010312.  
FT CONFLICT 229 T -> P (In REF\_5).  
SQ SEQUENCE 377 AA; 41527 MW; 7ECAB0FOC7F4143B CRC64;  
REVISION TO 24.

Query Match 7.1%; Score 115.5; DB 1; Length 377;  
Best Local Similarity 21.3%; Pred. No. 0.027; Gaps 15;  
Matches 69; Conservative 43; Mismatches 117; Indels 95;

QY 19 PAPAGAAKOKKV--EPPNAFGV-AAPPLPQASRLQAKRDPSPVSGPV----- 62  
Db 89 PARAGTYCGRGEPHPSPWAPSAPSPVIVMVGLEYPSLARPGRPVRTGENVTLSRSS 148  
QY 149 QSSFDIYHSR-EGRAHRLPAVPSINGTQADF-PLGPAT-HGTYR--CFGSHGSP 203  
QY 109 HEWSIANNTFTGMKMRDODACRSRSROSKVELAGKSNRARVSEESTCVVALLTEPLV 168  
Db 204 YEWSDASDPL-----PVSTGNSSSSWSPTERS-----FTGIA 238  
QY 222 PEENEPTELEGGSDFSTETLENCKRAHKSKEIRKIGLUTQHGIPTYRP-TETSNLE 280  
Db 292 -----EODFQBVTYAQLDHCITFORKTGTGSPORSK-----RPSTDNSVCI 311  
QY 281 HIGHEPTRAKSPFQLRGPGIGS 304  
Db 332 ELPNPNEAPALSPAHHEHHSQALMGS 355

RESULT 5  
VP2\_CAVC1 ID VP2\_CAVC1 STANDARD PRT 216 AA.  
ID\_Q93151; P54051; DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Protein VP2.  
OS Chicken anemia virus (German isolate Cuhavener-1) (CAV), and  
OS chicken anemia virus (USA isolate CFA-1) (CAV).  
OC Viruses; ssDNA viruses; Circoviridae; Gyrovirius.  
OX NCBI\_TaxID:73475, 73478,  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Cuhavener-1;  
RX MEDLINE:91237831; PubMed=1851873;  
RA Noteborn M.H.M., de Boer G.F., van Roozelaar D.J., Karreman C.,  
RA Kranenburg O., Vos J.G., Jeurissen S.H.M., Hoeben R.C., Zantema A.,  
RA Koch G., van Ormondt H., van der B. A.J.;  
RT "Characterization of cloned chicken anemia virus DNA that contains  
RT all elements for the infectious replication cycle.",  
RT J. Virol. 65:3131-3139(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Cuhavener-1;  
RX MEDLINE:92286898; Pubmed=1605740;  
RA Meekan B.M., Todd D., Creelan J.L., Barle J.A.P., Hoey E.M.,  
RA McNulty M.S.;  
RT "Characterization of viral DNAs from cells infected with chicken  
RT anemia agent: Sequence analysis of the cloned replicative form and  
RT transfection capabilities of cloned genome fragments.",  
RT Arch. Virol. 124:301-319(1992).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CIA-1;  
RX MEDLINE:97126092; PubMed=8971016;  
RA Renshaw R.W., Soine C., Weinkle T., O'Connell P.H., Ohashi K.,  
RA Watson S., Lucio B., Harrington S., Schat K.A.,  
RT "A hypervariable region in VPI of chicken infectious anemia virus  
RT mediates rate of spread and cell tropism in tissue culture.",  
RT J. Virol. 70:8872-8878(1996).  
RN [4]  
REVISION TO 24.

```

RA Renshaw R.W.; CC
RL Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases. CC
CC - - - - - FUNCTION: MAY BE A NON-STRUCTURAL PROTEIN THAT AT SOME STAGE OF CC
CC INFECTION IS REQUIRED FOR VIRUS ASSEMBLY. CC
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration CC
CC between the Swiss Institute of Bioinformatics and the EMBL outstation - CC
CC the European Bioinformatics Institute. There are no restrictions on its use CC
CC by non-profit institutions as long as its content is in no way CC
CC modified and this statement is not removed. Usage by and for commercial/ CC
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch). CC
CC
CC DR EMBLIB; M55918; AAA91822.1; -
CC DR EMBLIB; M81223; AAA42882.1; -
CC DR EMBLIB; L14767; AADD9422.1; -
CC PIR; A3926; A3926;
CC DR PIR; A4834; A4834;
CC DR PIR; A4834; A4834;
CC DR Interpro; IPR06945; Circovirus VP2.
CC FT PFM; PF04861; Circovirus VP2; 1.
CC FT CONFLICT .153 153 V -> A (IN REF. 1).
CC FT CONFLICT .187 187 D -> N (IN REF. 2).
CC SEQUENCE 216 AA; 24138 MW; F82699P88BDDE52 CRC64;
CC
CC Query Match 5.8%; Score 95.5; DB 1; Length 216;
CC Best Local Similarity 19.6%; Pred. No. 0.65; Gaps 14;
CC Matches 54; Conservative 44; Mismatches 74; Indels 103; Gaps 14;
CC
CC OY 17 GGPAPAGAAKMKVVEEPNAFGVNNPFLPQASRILQAKRDPSPVSPGVPHLFRILSGKCPSLVE 76
CC Db 21 QGPGFSGARQGQVINN-----ERSPR-----RSVTRITNGQ 52
CC
CC OY 77 STYXKEFCPFPNVHQECPFRMNAWSGLIGW-----HEWEIAN-NITGMAMRDGDAC 129
CC Db 53 ATNKFTAVENPSLQDPMWVN-YNHSAWLRBCSRSHAKICNCGQFRKHWFQACGL 111
CC
CC QY 130 RSPRSRKSLBLACKSNRLA-----HVSPTCIVYALTFETPVLCHHLLWPT 179
CC Db 112 EDQRSTQSLEEAFLRPLVQGRKAKRQKLDYHYSQPT-----PARKVKYKT 156
CC
CC QY 180 LPEALQRQDQEQLADLAEITPOGHKLURTLFEDAGYIKTPRENEPTEQLEGGSIDGF 239
CC Db 157 V-----RN-----QDEIAD-----READF--TSFEE----DGGTTSSDP 184
CC
CC QY 240 ETLENCRKAHKELSKERIKRKGILQ-HGIPYTRP 273
CC
CC Db 185 D-----EDINFDIGGDSGIVDELLGRPFPP 210
CC
CC
CC RESULT 6
CC VP2_CAV82 STANDARD; PRT; 216 AA.
CC ID VP2_CAV82
CC AC P54093;
CC DT 01-OCT-1996 (Rel. 34, Created)
CC DT 01-OCT-1996 (Rel. 34, Last sequence update)
CC DE 01-OCT-1996 (Rel. 34, Last annotation update)
CC OS Protein VP2.
CC OC chicken anemia virus (Japanese isolate 82-2) (CAV).
CC OX NEBILI-TAXID=73476;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RX MEDLINE=95297149; PubMed=7778281;
CC RA Kato A., Fujino M., Nakamura T., Ishihama A., Otaki Y. ;
CC RT "Gene organization of chicken anemia virus." ;
CC RL Virology 203:480-488 (1995).
CC -1- FUNCTION: MAY BE A NON-STRUCTURAL PROTEIN THAT AT SOME STAGE OF
CC INFECTION IS REQUIRED FOR VIRUS ASSEMBLY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its use
CC by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial/
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
```

CC modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/>) or send an email to license@isb-sib.ch.

--> EMBL: D31965; BAA06732;1; -

--> InterPro: IPR006945; Circovirus VP2.

--> Pfam: PF04861; Circovirus VP2.

--> DR

--> SEQUENCE 216 AA; 24138 MW; E23169F592BDDDE52 CRC64;

Query Match 5 8%; Score 94 5; DB 1; Length 216;  
Best Local Similarity 19 6%; Pred. No. 0. 78; Mismatches 74; Indels 103; Gaps 14;  
Matches 54; Conservative 44; Mismatches 74; Indels 103; Gaps 14;

OY 17 GGPARGRAAKKVVTEPPNAGVNNPFPQASRQKRPSPVSPVHLRLSKCFSIVE 76  
Db 21 GQPGPSGGAQGQVLSN-----ESPR-----RSPTINGVQ 52

OY 77 STYKFFCPFRNVTQHBEQTFRWNAYSGILGW----HEWETAN-NFTGMMNRDDAC 129  
Db 53 ATNKTFAVGVSQIQRDPDWYRN YHSHSAWWRRECSASHAKCNCQFORKWQAEGL 111

OY 130 RSRSRSKVILAGSNSRLA-----HVSEPTCTVYALIFETPLVCHPRLPHPT 179  
Db 112 EDRTSQAISLBEAIRPLRVLVGKRAKRKLDHYSQPT-----PAKKVKT 156

OY 180 LPEAQQRQWQEQVADLITPQCHERKURTFEDAGYKLTPEENERTOLEGPDGDF 239  
Db 157 V-----RW---KDLAD-----READF--TPSEE----DGTTSSDF 184

OY 240 ETLENCRKAHKELSKIKRKLGLTQ-HGIPYTRP 273  
Db 185 D-----EDINFDGGSGIVDELLSRPFTP 210

RESULT 7

VP2_CAV26	STANDARD;	PRT;	216 AA.
ID VP2_CAV26			
AC P54022;			
DT 01-OCT-1996 (Rel. 34, Created)			
DT 01-OCT-1996 (Rel. 34, Last sequence update)			
DT 01-OCT-1996 (Rel. 34, Last annotation update)			
DB Protein VP22.			
OS Chicken anemia virus (USA isolate 26p4) (CAV)			
OC Viruses; ssDNA viruses; Circoviridae; Gyrovirinae.			
OX NCBI_TAXID=73477;			
RN [1] SEQUENCE FROM N.A.			
RP MEDLINE=9134190; PubMed=1908516;			
RA Claesens J.A.J., Schrier C.C., Mackett A.P.A., Jagt E.H.J.M.,			
RA Sondermeijer P.J.A.;			
RT "Molecular cloning and sequence analysis of the genome of chicken			
RT anaemia agent.", 72:2003-2006(1991).			
RL J. Gen. Virol. 72:2003-2006(1991).			
CC INFECTON IS REQUIRED FOR VIRUS ASSEMBLY.			
CC			
CC This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC the European Bioinformatics Institute. There are no restrictions on its			
CC use by non-profit institutions as long as its content is in no way			
CC modified and this statement is not removed. Usage by and for commercial			
CC entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> )			
CC or send an email to license@isb-sib.ch).			
CC			
DR EMBL: D1008; BAA00959;1; -			
DR InterPro: IPR006945; Circovirus VP2.			
DR Pfam: PF04861; Circovirus VP2; T-			
DR Sequence: 216 AA; 24066 MW; CD4B69F88BDDC72 CRC64;			
SQ			

QY 17 GGPAPAGAKMVKVEPNAFGVNPFLQASRLOAKRSPSPVSGPPVHLRLSLCKSCFSLVE 76  
Db 21 GQPQPSGAQAOQVISN-----ERSPR-----RSRTINGVQ 52  
QY 77 STYKKECFPFHNTQHEQFRMAYSGLIGW---HEWEIAN-NFTGMMRDGDAC 129  
Db 53 ATTKFTAVGNPSIQRDRPYWRNN-YNNSIAWNRCSRSRSHAKCNCDFRKWFOQK 111  
QY 130 RSPRSRQSKVELAGCKSNRLA-----HVSSEPTCVYALTFPLVCHHALLVPT 179  
Db 112 EDRSTQASLEALIRPLRVQGRKAQKLDYHSQPT-----PKRKYKT 156  
QY 180 LPFALIQRMWDQEQDQLABLITPOGEKHLURTFEDACYLKTPEENEPETOQLEGPDLSLG 239  
Db 157 V-----RW----QDELAD-----READF--TPSE----DGGTSSDF 184  
QY 240 ETLENCRKAKELSKERIKLKGILT-HGIPYRP 273  
Db 185 D-----GDINPDIQGDSGTWELIGRPFTP 210

RESULT 8  
INVO\_AOTTR STANDARD; PRT; 544 AA.  
ID INVO\_AOTTR P24708;  
AC 01-MAR-1992 (Rel. 21, Created)  
DT 01-MAR-1992 (Rel. 21, Last sequence update)  
DE Involutin.  
GN Aotus trivirgatus (Night monkey) (Douroucouli).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Aotinae; Aotus.  
OX NCBI\_TAXID=9505;  
RN [1]  
SP SOURCE FROM N.A.  
RN MEDLINE=90014142; PubMed=2507864;  
RN RA Tseng H.; Green H.;  
RN "The involutin gene of the owl monkey: origin of the early region.";  
RN Mol. Biol. Evol. 6:460-468(1989).  
CC -!- FUNCTION: Part of the insoluble cornified cell envelope (CE) of  
stratified squamous epithelia.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic. Constituent of the scaffolding  
of the cornified envelope.  
CC -!- TISSUE SPECIFICITY: Keratinocytes of epidermis and other  
stratified squamous epithelia.  
CC -!- PTM: Substrate of transglutaminase. Specific glutamines or lysines  
are cross-linked to keratins, desmoplakin and to inter involucrin  
molecules.  
CC -!- SIMILARITY: Belongs to the involutin family.

This SWISS-PROT entry is copyright. It is produced through a collaboration - between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; M25313; AAH35375; 1; -.  
DR InterPro; IPR02360; Involutin.  
DR Pfam; PF00904; Involutin; 39.  
DR PROSITE; PS00795; INVOLUCRIN; 1.  
DR Keratinization; Repeat.  
SQ SEQUENCE 544 AA; 63927 MN; 2A02ABA5E1499F9D CRC64;

Query Match . 5.6%; Score 92; DB 1; Length 544;  
Best Local Similarity 23.4%; Pred. No. 4.2;  
Matches 34; Conservative 25; Mismatches 48; Indels 38; Gaps 4;

RESULT 9  
MLL2\_HUMAN STANDARD; PRT; 5262 AA.  
ID MLL2\_HUMAN P04686; O14687;  
AC 014686; O14687;  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DT 11-MAR-2004 (Rel. 43, Last annotation update)  
DE Myeloid/Lymphoid or mixed-lineage leukemia protein 2 (AML-related  
DB protein).  
RN MLL2 OR ALR.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
SP SOURCE FROM N.A. (ISOFORMS 1; 2 AND 3).  
RN MEDLINE=97388474; PubMed=247308;  
RN Prasad R., Zhadanov A.B., Sedkov Y., Bulirich F., Druck T., Mazo A.,  
RA Rallapalli R., Yano T., Alder H., Croce C.M., Huebner K., Mazo A.,  
RA Canaani E.;  
RT "Structure and expression pattern of human ALR, a novel gene with  
RT strong homology to ALU-1 involved in acute leukemia and to Drosophila  
RT trithorax.,"  
RL Oncogene 15:549-560(1997).  
RN [2]  
SP INTERACTION WITH ASC-2/NCOA6 CONTAINING COMPLEX.  
RN TISSUE-Cervical carcinoma;  
RN MEDLINE=22371456; PubMed=12482668;  
RN Goo Y.-H., Sohn Y.-C., Kim D.-H., Kim S.-W., Kang M.-J., Jung D.-J.,  
RA Kwak E., Barley N.A., Berger S.L., Chow V.T., Roeder R.G.,  
RA Azorsa D.O., Meltzer P.S., Suh P.-G., Song E.J., Lee X.-J., Lee Y.C.,  
RA Lee J.W.;  
RT "Activating signal cointegrator 2 belongs to a novel steady-state  
RT complex that contains a subset of trithorax group proteins.,"  
RL Mol. Cell. Biol. 23:140-149(2003).  
CC -!- FUNCTION: May be involved in transcriptional regulation.  
CC -!- SUBUNIT: Belongs to the ASC-2/NCOA6 complex (Ascom), which  
contains ASC-2/NCOA6, the retinoblastoma-binding protein RBQ-3/  
RBQ5, alpha-, beta- and gamma-tubulins, the trithorax group proteins  
MLL2 and MLL3, and ASH2/ASC12.  
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -!- ALTERNATIVE PRODUCTS:  
Event=Alternative splicing; Named isoforms=3;  
Name=1; IsoId=014686-1; Sequence=Displayed;  
Name=2; IsoId=014686-2; Sequence=VSP\_008563, VSP\_008559;  
Name=3; IsoId=014686-3; Sequence=VSP\_008560;  
CC -!- TISSUE SPECIFICITY: Expressed in most adult tissues, including a  
variety of hematopoietic cells, with the exception of the liver.  
CC -!- MISCELLANEOUS: This gene mapped to a chromosomal region involved  
in duplications and translocations associated with cancer.  
CC -!- SIMILARITY: Belongs to the transcription factor trithorax family.  
CC -!- SIMILARITY: Contains 5 PHD-type zinc fingers.  
CC -!- SIMILARITY: Contains 1 post-SET domain.  
CC -!- SIMILARITY: Contains 1 RING-type zinc finger.  
CC -!- SIMILARITY: Contains 1 SET domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration - between the Swiss Institute of Bioinformatics and the EMBL outstation -





RC TISSUE=Kidney;  
 RX MEDLINE=96370811; PubMed=8774707;  
 RA Riento K., Jaenti J., Jansson S., Hielm S., Lehtonen E., Ehnlholm C.,  
 RA Keränen S., Oikonen V.M.;  
 RT "A sec1-related vesicle-transport protein that is expressed  
 predominantly in epithelial cells";  
 RL Bur. J. Biochem. 239(638-646)(1996).  
 CC -!- FUNCTION: Involved in the protein trafficking from the Golgi  
 apparatus to the plasma membrane (By similarity).  
 CC -!- SIMILARITY: Binds syntaxins 1A, 2, 3 but not 4 (By similarity).  
 CC -!- SIMILARITY: Belongs to the Syntaxin-binding protein family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to license@isb-sib.ch).  
 CC  
 CC SEQUENCE FROM N.A.  
 DR Protein transport; Sec1; 1.  
 DR Pfam; PF00995; Sec1; 1.  
 DR InterPro; IPR001619; Sec1-like.  
 DR EMBL; L41609; AAC00031.1; --.  
 DR PIR; P33176; T51930.  
 DR HSPB; P33176; T51930.  
 DR InterPro; IPR007420; DUF455.  
 DR PIR; T51930; T51930.  
 DR Pfam; PP0425; DUF465; 1.  
 DR PRINTS; PRO0380; KINESIN.  
 DR SMART; SM00129; KINESIN.  
 DR PROSITE; PS00411; KINESIN\_MOTOR\_DOMAIN1; 1.  
 DR PROSITE; PS0067; KINESIN\_MOTOR\_DOMAIN2; 1.  
 DR KW Motor protein; Microtubule; ATP-binding; Coiled coil.  
 FT DOMAIN 1 341 KINESIN-MOTOR (BY SIMILARITY).  
 FT DOMAIN 2 887 894 COILED COIL (POTENTIAL).  
 FT DOMAIN 3 888 934 GLOBULAR.  
 FT NP BIND 87 94 ATP (POTENTIAL).  
 FT NT-BIND 237 244 ATP (POTENTIAL).  
 SQ FCFP 86  
 254 YDLINEEQDTTRYETGGLSEAREKAVLIDDDLWELVRLRHIAVDSSKKVTELLKFCES 313  
 Db QY 87 HMTVQHQTRFWNAYSGIIGIWHMELIANTFTGMMRDRDADCRSRSSRSQVKELAGKSN 146  
 Db 314 KRITDPAKNT--DLSHIKKMQPKQYKLKNS--THLHADCMKFKKS--VEKLKGHQ 369  
 QY 147 RLAHVSPSTCVIALTPTPLVCHPHALLVPTPLPEALQRQDW-----P 192  
 Db 370 DLAMGSDT-----EKGKIKDANKLIV-PVLDAAVYAKDVRKVLVILLRGVSE 419  
 QY 193 QDLADELI--TPOQHBLKLRLTFEDAGYLUKPTPEENPTOLEGGP 234  
 Db 420 ENLA-XLIQHANQAHSSLRNLQDQLGTVNP-----GGP 454  
 RESULT 12  
 KINH SYNA STANDARD; PRT; 935 AA.  
 ID AC 043093; 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Kinesin heavy chain (Synkin).  
 OS Syncphaelastrum racemosum.  
 OC Baryota; Fungi; Zygomycota; Zygomycetes; Mucorales;  
 OC Syncphaelastaceae; Syncphaelastrum.  
 OX NCBI\_TAXID=13706;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9627360; PubMed=9613604;  
 RA Grummt M., Pistor S., Lottepeach F., Schliwa M.;  
 RT "Cloning and functional expression of a 'fast' fungal kinesin.";  
 RL FEBS Lett. 427:79-84(1998).  
 CC -!- FUNCTION: KINESIN IS A MICROTUBULE-ASSOCIATED FORCE-PRODUCING  
 PROTEIN THAT MAY PLAY A ROLE IN ORGANELLE TRANSPORT. ITS MOTOR  
 ACTIVITY IS DIRECTED TOWARD THE MICROTUBULE'S PLUS END. THE SPEED  
 OF THIS MOTOR IS 4-5 TIMES FASTER THAN ITS ANIMAL COUNTERPARTS.  
 CC  
 CC -!- DOMAIN: Composed of three structural domains: a large globular N-  
 terminal domain which is responsible for the motor activity of  
 kinesin (it hydrolyzes ATP and binds microtubule), a central  
 alpha-helical coiled coil domain that mediates the heavy chain  
 dimerization; and a small globular C-terminal domain which  
 interacts with other proteins (such as the kinesin light chains),  
 vesicles and membranous organelles.  
 CC -!- SIMILARITY: Belongs to the kinesin-like protein family. Kinesin  
 subfamily.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to license@isb-sib.ch).  
 CC  
 DR PIR; T51930; T51930.  
 DR HSPB; P33176; T51930.  
 DR InterPro; IPR007420; DUF455.  
 DR PIR; T51930; T51930.  
 DR Pfam; PP0425; DUF465; 1.  
 DR PRINTS; PRO0380; KINESIN.  
 DR SMART; SM00129; KINESIN.  
 DR PROSITE; PS00411; KINESIN\_MOTOR\_DOMAIN1; 1.  
 DR PROSITE; PS0067; KINESIN\_MOTOR\_DOMAIN2; 1.  
 DR KW Motor protein; Microtubule; ATP-binding; Coiled coil.  
 FT DOMAIN 1 341 KINESIN-MOTOR (BY SIMILARITY).  
 FT DOMAIN 2 887 894 COILED COIL (POTENTIAL).  
 FT DOMAIN 3 888 934 GLOBULAR.  
 FT NP BIND 87 94 ATP (POTENTIAL).  
 FT NT-BIND 237 244 ATP (POTENTIAL).  
 SQ FCFP 86  
 254 YDLINEEQDTTRYETGGLSEAREKAVLIDDDLWELVRLRHIAVDSSKKVTELLKFCES 313  
 Db QY 87 HMTVQHQTRFWNAYSGIIGIWHMELIANTFTGMMRDRDADCRSRSSRSQVKELAGKSN 146  
 Db 314 KRITDPAKNT--DLSHIKKMQPKQYKLKNS--THLHADCMKFKKS--VEKLKGHQ 369  
 QY 147 RLAHVSPSTCVIALTPTPLVCHPHALLVPTPLPEALQRQDW-----P 192  
 Db 370 DLAMGSDT-----EKGKIKDANKLIV-PVLDAAVYAKDVRKVLVILLRGVSE 419  
 QY 193 QDLADELI--TPOQHBLKLRLTFEDAGYLUKPTPEENPTOLEGGP 234  
 Db 420 ENLA-XLIQHANQAHSSLRNLQDQLGTVNP-----GGP 454  
 RESULT 13  
 YM9\_YEAST STANDARD; PRT; 1328 AA.  
 ID AC 003434; 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Transposon Tyl protein B.  
 GN T1B OR YML039W OR YMB054.04.  
 OS Saccharomyces cerevisiae (Baker's Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TAXID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=S288C / AB972;  
 MEDLINE=9713268; PubMed=9169872;  
 RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,  
 RA Connor R., Deciman K., Devlin K., Gentles S., Hamlin N., Hunt S.,  
 RA Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,  
 RA Rice P., Shelton J., Walsh S., Whitehead S., Barrell B.G.;  
 RT "The nucleotide sequence of *Saccharomyces cerevisiae* chromosome  
 XIII.";



239	--NMVAKCGRYATPRRDGHILIGSTLEHSQFDKPTD-----BAQESLRA	280	DR PROSITE; PS00886; TRBD; 1.
Qy			KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; AMP-binding;
Db	248 AHKELSKERKLK---GIL-TQHGPYTRP 273 281 SAEELIPELADMOPVAHAGLRLPGSPEGIPIGP 314		KW RNA-binding; tRNA-binding; Metal-binding; Zinc; Complete proteome.
			FT SITE 15 "HIGH" REGION.
			FT DOMAIN 25 "KNSKS" REGION.
			FT METAL 703 TRNA-BINDING.
			FT METAL 147 ZINC (BY SIMILARITY).
			FT METAL 150 ZINC (BY SIMILARITY).
			FT METAL 160 ZINC (BY SIMILARITY).
			FT METAL 163 ZINC (BY SIMILARITY).
			FT BINDING 348 ATP (BY SIMILARITY).
			FT BINDING 348 ATP (BY SIMILARITY).
			FT STRAIN=TLS / ATCC 49632 / DSM 12025;
			FT MEDIUM=2210385; PubMed=1203901;
			RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,
			RA Dodson R.J., Deboy R., Gwinn M.L., Nelson W.C., Haft D.H.,
			RA Hickie E.K., Peterson J.D., Durkin S.S., Kolonay J.L., Yang F.,
			RA Holt I., Umamaheshwaran L.A., Mason T., Brenner M., Shea T.P., Parksey D.,
			RA Vamathevan J.C., Feldblyum T.V., Hansen C.L., Craven M.B., Radune D.,
			RA Ventter J.C., Khouri H., White O., Gruber T.M., Ketchum K.A.,
			RT "The complete genome sequence of Chlorobium tepidum TLS, a
			RT photosynthetic, anaerobic, green-sulfur bacterium.",
			Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).
			CC i- FUNCTION: Is required not only for elongation of protein synthesis
			but also for the initiation of all mRNA translation through
			initiator tRNA(fMet) aminoacylation (By similarity).
			i- CATALYTIC ACTIVITY: ATP + L-methionine + tRNA(Met) = AMP +
			diphosphate + L-methionyl-tRNA(Met).
			CC i- COFACTOR: Binds 1 zinc ion per subunit. (By similarity).
			CC i- SUBUNIT: Homodimer (By similarity).
			CC i- SUBCELLULAR LOCATION: Cytoplasmic.
			-i- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
			-i- SIMILARITY: MetG subfamily 1.
			CC This SWISS-PROT entry is copyright. It is produced through a collaboration
			between the Swiss Institute of Bioinformatics and the EMBL outstation -
			the European Bioinformatics Institute. There are no restrictions on its
			use by non-profit institutions as long as its content is in no way
			modified and this statement is not removed. Usage by and for commercial
			entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>
			or send an email to license@isb-sib.ch).
			CC
			DR EMBL; AE018861; AAM72204.1; -.
			DR TIGR; CT0969; -.
			DR HAMAP; MF_00038; fused; 1.
			DR InterPro; IPR004195; MetG Cterm.
			DR InterPro; IPR00324; MetRS dimerising.
			DR InterPro; IPR008994; Nucleic acid_DNA.
			DR InterPro; IPR002300; tRNA_synt_1a.
			DR InterPro; IPR001412; tRNA-synt_1.
			DR InterPro; IPR002304; tRNA_synt_m.
			DR InterPro; IPR002547; tRNA_bind.
			DR Pfam; PF00133; tRNA_synt_1; 1.
			DR Pfam; PF01588; tRNA_bind; 1.
			DR PIRSP001528; MetRS dimerising; 1.
			DR PRINTS; PRO041; tRNA-SYNTNET.
			TIGRFAMS; TIGR00398; metG; 1.
			PROTEINS; PS00178; AA_TRNA_DIGASE_I; 1.

This Page Blank (uspto)

Copyright (c) 1993 - 2004 Compugen Ltd.

GenCore version 5.1.6

Om protein - protein search, using sw model

Run on: July 26, 2004, 11:05:03 ; Search time 19.9295 Seconds  
(without alignments)

4828.665 Million cell updates/sec

Title: US-10-657-280-3

Perfect score: 1634

Sequence: 1 MAAGLARILLLGLISAGGSPAA.....TPRAKSPEQLRQDPGLRGSL 305

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 31518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL 25:\*

1: sp\_archaea:\*

2: sp\_bacteriad:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phage:\*

10: sp\_plant:\*

11: sp Rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_rvirus:\*

16: sp\_bacteriapl:\*

17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

**SUMMARIES**

Result No.	Score	Query Match Length	DB ID	Description
1	1634	100.0	305	4 Q9UJ9
2	1618.5	99.1	304	4 Q96L13
3	1601	98.0	331	4 Q96RZ2
4	1301.5	79.7	307	11 Q7NEO
5	1288	78.8	314	11 Q8CSJ3
6	185.5	11.4	507	5 Q3440
7	172.5	10.6	529	13 Q802Z2
8	171	10.5	547	10 Q9FM6
9	171	10.5	647	10 Q8LDD0
10	155.5	9.5	612	4 Q9BW99
11	155.5	9.5	615	4 Q81Z58
12	148	9.1	548	5 Q9VJD1
13	141	8.6	521	11 Q08795
14	141	8.6	528	11 Q921X2
15	140.5	8.6	200	4 Q9EBU9
16	140.5	8.6	398	4 Q96D06

**ALIGNMENTS**

RESULT	ID	DESCRIPTION	PRT;	305 AA.
1	Q9UJ9	PRELIMINARY;		
AC	Q9UJ9;			
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)			
DE	C316G12.3 (CAB56184).			
GN	Homo sapiens (Human).			
OS	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OC	NCBI_TaxID=9606;			
OX	[1]			
RN	SEQUENCE FROM N.A.			
RA	Frankland J.;			
RL	Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE_Brain;			
RA	Fitzgerald P., Amarante-Mendes G.P., Li W., Green D.R.;			
RT	"CDNA from human fetal brain.";			
RL	Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AL031709; CAB56184.1; -.			
DR	EMBL; AR302186; AAG27706.1; -.			
DR	EMBL; TA5062; T45562.			
DR	InterPro; IPR009011; Man_6_P_R_bind.			
DR	SEQUENCE: 305 AA; 33973 MW; -7774BBC0911DA1C2 CRC64;			
QC	Query Match 100.0%; score 1634; DB 4; length 305; Best Local Similarity 100.0%; Pred. No. 1.2e-148; Indels 0; Gaps 0;			
Db	1 MAAGLARILLLGLISAGGSPAPAGAKKVKVEEPNAFVNPNFLPQASSLQAKRDPSPVSG			
Db	1 MAAGLARILLLGLISAGGSPAPAGAKKVKVEEPNAFVNPNFLPQASSLQAKRDPSPVSG			
QY				
QY	61 PVHIFRLSKCFFSLVESTYKVFCPFPNVTOHQTFERNWAVNAYGILGJWHENBIANTFTG			
QY	120			



RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Maruyama K., Farmer A.A., Rubin G.M., Hong II.,  
 RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carminci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., Mcowan P.J., McKernan K.J., Malek J.A., Gunnarino P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzzey D.M., Sodergren E.J., Li X., Gibbs R.A.,  
 RA Farney J., Helton E., Kettman J.A., Nadan A., Bouffard G.G.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Rodriguez M.C.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Smalius D.E., Schnurch A., Schein J.E.,  
 RA Jones S.J., Maxx M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences";  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 DR [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FVB/N; TISSUE=Salivary gland;  
 RA STRAUBERG R.; Submitted (AUG-2003) to the EMBL/GenBank/DDBJ databases.  
 RL EMBL; BC05872; AAC5872.1; -.  
 DR Hypothetical protein;  
 SQ SEQUENCE 307 AA; 8712A9E4288D556 CRC64;

Query Match 79.7%; Score 1301.5; DB 11; Length 307;  
 Best Local Similarity 78.8%; Pred. No. 1.2e-116;  
 Matches 241; Conservative 25; Mismatches 39; Indels 1; Gaps 1;

QY 1 MAGIAGRLLILGLISAGGPAGAPAGAKKVKVTPENPAFGVNNPFLPOASRQIAKRPSPSG 60  
 Db 1 MAGRIAGTLMULGLAQGPPACAGKVKVTPENPFLQNLPLQASRQPKRPSAVG 60  
 QY 61 PVHLFRISGKCSLIVESTYKVEFPCPFHNTQHEQPRWNAYSGIGIWHWEIANNFTG 120  
 Db 61 PLHLFRAGKCSLIVESTYKVEFPCPFHNTQHEQPRWNAYSGIGIWHWEIANNFTG 120  
 QY 121 MNWRDGDACRSRSSRSQKVLACGKSNLAVYSEPSITCVIALTFPLVCHPHALIYVPTI 180  
 Db 121 MNWTDGDSCHSRSSRSQKPLRSLQSVELTGKINLRKAVSEPSITCVIALTFPLVCHPHS 180  
 Db 121 MNWTDGDSCHSRSSRSQKPLRSLQSVELTGKINLRKAVSEPSITCVIALTFPLVCHPHS 180  
 QY 181 PEALQRQWDQEQDADELITPQGHKLRLTPEGDYKLTPEEMEPTEQEGDLSIGE 240  
 Db 181 SEALQWMDQEQDADELITPQGHKLRLTPEGDYKLTPEEMEPTEQEGDLSIGE 240  
 QY 241 TLENCRKAHKELSKETKRLKGILTQHAIPIYRPTETSNLIEHGHBT-RAKSPEOLRGDP 299  
 Db 241 TLDNCRKAHAELOSOEVRLTSJLQHQHGIPTQTTTHSOSHQQLPIGAIAEHRSDP 300  
 QY 300 GLR GSL 305  
 Db 301 GLRGNI 306

RESULT 5

Q8C5J3 PRELIMINARY; PRT; 314 AA.

ID Q8C5J3; PRELIMINARY; PRT; 314 AA.

AC Q8C5J3; (Created)  
 DT 01-MAR-2003 (TREMBrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBrel. 23, Last annotation update)

DE Hypothetical Extracytoplasmic domain of cation-dependent mannose 6-phosphate receptor structure containing protein.  
 DE GN 130000G08RJK.

OS Mus musculus (Mouse);  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. NCBI\_TaxID=10090;

[1] SEQUENCE FROM N.A. TISSUE=Olfactory brain;  
 RP STRAIN=CC-7BL/6J; TISSUE=Olfactory brain;  
 RC MEDLINE=22354683; PubMed=12466851;

RA The FANTOM Consortium;  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RA "Analysis of the mouse transcriptome based on functional annotation of  
 RA 60,770 full-length cDNAs";  
 RA Nature 420:563-573 (2002).  
 RA EMBL; AK078330; BAC37183.1; -  
 DR MGII: MGII:1921404; 1>0004008Rik.  
 DR InterPro; IPR00011; Man\_6\_P\_R\_bind.  
 DR KW Hypothetical protein.  
 SQ SEQUENCE 314 AA; 34891 MW; 7C3424A15649D058 CRC64;

Query Match 78.8%; Score 1288; DB 11; Length 314;  
 Best Local Similarity 77.0%; Pred. No. 2.4e-115;  
 Matches 241; Conservative 25; Mismatches 39; Indels 8; Gaps 2;

QY 1 MAGIAGRLLILGLISAGGPAGAPAGAKKVKVTPENPAFGVNNPFLQASRQIAKRPSPSG 60  
 Db 1 MAGRIAGTLMULGLAQGPPACAGKVKVTPENPFLQNLPLQASRQPKRPSAVG 60  
 QY 61 PVHLFRISGKCSLIVESTYKVEFPCPFHNTQHEQPRWNAYSGIGIWHWEIANNFTG 120  
 Db 61 PLHLFRAGKCSLIVESTYKVEFPCPFHNTQHEQPRWNAYSGIGIWHWEIANNFTG 120  
 QY 121 MNWRDGDACRSRSSRSQKVLACGKSNLAVYSEPSITCVIALTFPLVCHPHALIYVPTI 180  
 Db 121 MNWTDGDSCHSRSSRSQKPLRSLQSVELTGKINLRKAVSEPSITCVIALTFPLVCHPHS 180  
 QY 174 LUYPITPRLAQWQDVQEQLADELTPOGHKLRLTPEGDYKLTPEEMEPTEQEGD 233  
 Db 181 LLVYPTPSALQDQRWDQEVQEQLADELTPOGHKLRLTPEGDYKLTPEEMEPTEQEGD 240  
 QY 234 PDSLIGETTLENCRKAHKELSKETKRLKGILTQHAIPIYRPTETSNLIEHGHBT-RAKS 292  
 Db 241 SKIGIIGLETLNCRKAHKELSKETKRLKGILTQHAIPIYRPTETSNLIEHGHBT-RAKS 292  
 QY 293 EQLRGGPGRLGSL 305  
 Db 301 EHRSDPGLRGN 313

RESULT 6

Q23440 PRELIMINARY; PRT; 507 AA.

ID Q23440; Q22519; AC Q23440; Q22519; (Created)  
 DT 01-NOV-1996 (TREMBrel. 01, Created)  
 DT 01-MAY-1999 (TREMBrel. 10, Last sequence update)  
 DT 01-OCT-2003 (TREMBrel. 25, Last annotation update)  
 DE ZK1307.8 protein.  
 GS ZK1307.8.  
 OC Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderrinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RA Berks M.;  
 RA Submitted (TAN-1995) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; ZA7356; CA87420.1; - JOINED.  
 DR EMBL; 247358; CA87420.1; -.  
 DR EMBL; 247356; CA87438.1; JOINED.  
 DR PIR; T24944; T24344.  
 DR WormPep; ZKU1307.8; CB15547.  
 DR GO; GO:0005509; F:calcium ion binding; IEA.  
 DR InterPro; IPR002048; ER-hand.  
 DR InterPro; IPR000886; ER target; S.  
 DR InterPro; IPR002172; LDL receptor A.  
 DR InterPro; IPR009011; Man\_6\_P\_R\_bind.  
 DR Pfam; PF0036; lefhand; 1.  
 DR SMART; SM0192; India; 1.  
 DR PROSITE; PS00018; EF HAND; 2.  
 DR PROSITE; PS00014; ER TARGET; 1.  
 SQ SEQUENCE 507 AA; 58062 MW; 2764BC56741227DF CRC64;

Query Match 11.4%; Score 185.5; DB 5; Length 507;  
 Best Local Similarity 38.7%; Pred. No. 5.5e-09; Indels 7; Gaps 4;  
 Matches 41; Conservative 11; Mismatches 47; DR RL Sequence features of the regions of 1,456,315 bp covered by nineteen  
 Constitutive 11; Mismatches 47; Indels 7; Gaps 4;  
 DR physically assigned PI and TAC clones.";  
 DR DNA Res. 5.41-54(1998).  
 DR EMBI; AB00949; BAB1163.1; -.  
 DR InterPro; IPR000886; ER target\_S.  
 DR SMART; SM0192; LDDA; 1.  
 DR PROSITE; PS00014; ER TARGET; 1.  
 DR SEQUENCE 647 AA; 73213 MW; 49CC4F6DF7026472 CRC64;

RESULT 7

QY 67 LSGKOFSLVSTYKVERCPFHNTYQHGTFRNWAISGI-LGIWHENE-TANNTGMMWR 124  
 Db 394 LKOKCFDARVQQTQYQFPFGONTOKT---GAYSGTSLSFKENSGPEGSNKYRQHG 449  
 QY 125 DGDAC-RSRSROSKVELAGKSNSRLAHSYSPSTCVALTFRPLVC 169  
 Db 450 DGQOCWNQSPKRSDIDITBEGEENELVETPAKEYLFTFRPLAC 495

QY 30 VEERNAFGNNPFLPQASRL-----QAKRDPSPVSPGVHPLRSGRCFIVES 77  
 ID Q80222 PRELIMINARY; PRT; 529 AA.  
 DR Q80222; TYKVERCPFHNTYQHGTFRNWAISGI-LGIWHENE-TANNTGMMWR 124  
 AC 01-JUN-2003 (TREMBLrel. 24; Created)  
 DT 01-OCT-2003 (TREMBLrel. 25; Last annotation update)  
 DR 01-OCT-2003 (TREMBLrel. 25; Last annotation update)  
 DE Similar to protein kinase C substrate 80kH.  
 OS Brachydanio rerio (Zebrafish) (Danio rerio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.  
 OC NCBI\_TaxID:7955;  
 RN [1] SEQUENCE FROM N.A.  
 RP Strausberg R.; Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
 RU EMBL; BC046883; RAH4683.1; -.  
 DR GO; GO\_0016301; P:kinase activity; IFA.  
 DR InterPro; IPR002448; EF-hard.  
 DR PROSITE; PS00018; EF\_HAND; 1.  
 KW Kinase  
 SQ 529 AA; 59584 MW; BLEAF134CBE92502 CRC64;

Query Match Best Local Similarity 36.2%; Pred. No. 1e-07; Length 529; DR Q81D0  
 Matches 38; Conservative 12; Mismatches 50; Indels 5; Gaps 3; DR Q81D0; TYKVERCPFHNTYQHGTFRNWAISGI-LGIWHENE-TANNTGMMWR 125  
 DR PROSITE; PS00018; EF\_HAND; 1.  
 KW Kinase  
 SQ 412 LYSQCYEVSTSEVYRCPFHNVTQHGTFRNWAISGI-LGIWHENE-TANNTGMMWR 125  
 DR 126 GDAC-RSRSQVKELAGKSNSRLAHSYSPSTCVALTFRPLVC 169  
 DR 469 GTGCWQGRNRSTVKGKETMLTSEPSRCYLMFPTPVC 513

RESULT 8

QY 67 LSGKOFSLVSTYKVERCPFHNTYQHGTFRNWAISGI-LGIWHENE-TANNTGMMWR 124  
 ID Q9FM96 PRELIMINARY; PRT; 647 AA.  
 AC 09FM96; DR Q81D0; TYKVERCPFHNTYQHGTFRNWAISGI-LGIWHENE-TANNTGMMWR 125  
 DT 01-MAR-2001 (TREMBLrel. 16; Created)  
 DT 01-OCT-2003 (TREMBLrel. 25; Last annotation update)  
 DR "Full-length messenger RNA sequences greatly improve genome annotation.";  
 DE Arabidopsis thaliana (Mouse-ear cress).  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Barycoda; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; euroids II; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxID:3702;  
 DR [1] SEQUENCE FROM N.A.  
 DR Hais B.-J., Volkovskiy N., Town C.D., Troukhman M., Alexandrov N., Feldmann K.A., Flavell R.B., White O., Salzberg S.L.; DR "Full-length messenger RNA sequences greatly improve genome annotation.";  
 RT Genomic Biol. 0:0-0 (2002).  
 RL [2] SEQUENCE FROM N.A.  
 RP Brover V., Troukhman M., Alexandrov N., Lu Y.-P., Flavell R., Feldmann K.; DR Q9FM96; DR Q81D0; TYKVERCPFHNTYQHGTFRNWAISGI-LGIWHENE-TANNTGMMWR 125  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY086076; AAC66282.1; -.  
 DR InterPro; IPR00886; ER target\_S.  
 DR PROSITE; PS00014; ER\_TARGET; 1.  
 KW Hypothetical protein.  
 SQ 647 AA; 73200 MW; 49C948C5F46B6D2 CRC64;

Query Match Best Local Similarity 28.0%; Pred. No. 1.9e-07; Length 647; DR Q81D0; TYKVERCPFHNTYQHGTFRNWAISGI-LGIWHENE-TANNTGMMWR 125  
 Matches 54; Conservative 24; Mismatches 77; Indels 38; Gaps 8; DR SEQUENCE FROM N.A.  
 RU SEQUENCE FROM N.A.  
 DR STRAIN-Columbia; DR Q81D0; TYKVERCPFHNTYQHGTFRNWAISGI-LGIWHENE-TANNTGMMWR 125  
 DR MEDLINE=982056; PubMed=9628582;  
 DR Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., Miyajima N., Tabata S.; DR Q81D0; TYKVERCPFHNTYQHGTFRNWAISGI-LGIWHENE-TANNTGMMWR 125  
 DR "Structural analysis of Arabidopsis thaliana chromosome 5. IV.

Db	536	KTYKVCAYKATQEE----GYSKTRLGWEDKFE---NSYQFMSVTNGEKCWNGPDRS	586	DE Hypothetical protein (Fragment).
Qy	136	SKVELACGKSKVLRHSEPSRCVYALTFEPFLVPHALLVYPTPEAQORQWMQVEDL	195	OS Homo sapiens (Human).
Db	587	IKVKLKGKLNBLMDVDBPSRCEYAILSTPARC-----LEDKULKELOQKLEKLMNOD-	639	OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Qy	196	ADELITPQGHKEKL 208		OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Db	640	-----KPQNHDPL 647		OX NCBI_TAXID=9606;
RESULT 10				
Q9BW99	PRELIMINARY;	PRT;	612 AA.	SEQUENCE FROM N.A.
ID Q9BW99				TISSUE-Muscle;
AC Q9BW99;				Strausberg R.;
DT 01-JUN-2001 (TREMBrel. 17, Created)				Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DT 01-JUN-2001 (TREMBrel. 17, Last sequence update)				EMBL; BC02313; AAH2513.1; -.
DE Similar to amplified in osteosarcoma.				InterPro; IPR000011; Man_5_P_R_bind.
OS Homo sapiens (Human).				KW Hypothetical protein
OC Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				FT HYPO-TER 1
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				SQ SEQUENCE 615 AA; 69641 MW; 757D625B6FDA7097 CRC64;
OX NCBI_TAXID=9606;				Query Match 9.5%; Score 155.5; DB 4; Length 615;
RN [1]				Best Local Similarity 21.7%; Pred. No. 5.5e-06; Matches 79; Conservat-
RP TISSUE-Muscle;				ive 47; Mismatches 121; Indels 117; Gaps 18;
RC STRAUSBERG R.;				Db 21 PAGAAKMKVVEENARVANPNLPOASRLQKRDSPVSPVHLFRLSGKCFSLVESTYK 80
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.				79 PAGAAKMKVVEENARVANPNLPOASRLQKRDSPVSPVHLFRLSGKCFSLVESTYK 80
DR EMBL; BC000532; AAH0532.1; -.				Db 81 YERCPPHRVHQ-----HEOTPRWNWVSGIIGWHEIANNTFTGMM 123
DR EMBL; BC007254; AAH07254.1; -.				Db 123 YERCPPHRVHQHEDSBIGKVLVLYQYQSADWDDETAKSQHRLKTHSQTYG-- 179
DR InterPro; IPR0059011; Man_6_P_R_bind.				Db 124 RDGDAC--RSRSRSQSKELACSK-----SNRAHVBPESTCVYALTFEPFLVCG----- 169
SQ SEQUENCE 612 AA; 69246 MW; 1FD06FP7B23101015 CRC64;				Db 180 -NGSKCDINGRPRAEYRFLCDGAGISGYDIDRVDPDEPLSCSYVLTTRTPRILCPHPILRP 238
Query Match 9.5%; Score 155.5; DB 4; Length 612;				Qy 170 ---HPHALLVYPTL-PEA---LQRQWDOVE-----QDLABEL-----ITPO- 203
Best Local Similarity 21.7%; Pred. No. 5.5e-06; Matches 79; Conservative 47; Mismatches 121; Indels 117; Gaps 18;				Db 239 PRSSAAPQKLTCHPSLQPEEYMAVQROADSKOYGDKITEBLQDLPGVWSETKSGVAPOK 298
Qy 21 PAGAAKMKVVEENARVANPNLPOASRLQKRDSPVSPVHLFRLSGKCFSLVESTYK 80				Qy 204 -----GHEKLRLTFLEDA-GYLKTP-EENEPT-----OLEGGDSLGEETL 242
Db 76 PAGAAKMKVVEENARVANPNLPOASRLQKRDSPVSPVHLFRLSGKCFSLVESTYK 80				Db 299 MAGASPTKDSKSDFWKMLNPEPDQPGGEVPAEQQPSAADSAGPNDFQNNQ 358
Qy 81 YERCPPHRVHQ-----HEOTPRWNWVSGIIGWHEIANNTFTGMM 123				Db 243 ENCKAKHKLSEKTRKGLTOHGIPYTRPTETSNLIEHIGHETP-----RAKSPBQL 295
Db 120 YERCYGRHIOQYHMEDESEIKGEVLYLGYYQSADFWDDETAKSQHRLKTHSQTYG-- 176				Db 359 VKVIRSPADLIRFIBELKG-GTKKGK-----NIGEQPVDAAEVPOREPEKE 406
Qy 124 RODAAC--RSRSRSQSKELACOK-----SNRAHVBPESTCVYALTFEPFLC----- 169				Qy 296 RGDP 299
Db 177 -NGSKCDINGRPRAEYRFLCDGAGISGYDIDRVDPDEPLSCSYVLTTRTPRILCPHPILRP 235				Db 407 RGDP 410
Qy 170 ---HPHALLVYPTL-PEA---LQRQWDOVE-----QDLABEL-----ITPO- 203				
Db 236 PRSSAAPQKLTCHPSLQPEEYMAVQROADSKOYGDKITEBLQDLPGVWSETKSGVAPOK 295				
Qy 204 -----GHEKLRLTFLEDA-GYLKTP-EENEPT-----OLEGGDSLGEETL 242				
Db 296 MAGASPTKDSKSDFWKMLNPEPDQPGGEVPAEQQPSAADSAGPNDFQNNQ 355				
Qy 243 ENCKAKHKLSEKTRKGLTOHGIPYTRPTETSNLIEHIGHETP-----RAKSPBQL 295				
Db 356 VKVIRSPADLIRFIBELKG-GTKKGK-----NIGEQPVDAAEVPOREPEKE 403				
Qy 296 RGDP 299				
Db 404 RGDP 407				
RESULT 11				
Q8IZ58	PRELIMINARY;	PRT;	615 AA.	SEQUENCE FROM N.A.
ID Q8IZ58				RP
AC Q8IZ58;				RC STRAIN-Berkeley;
DT 01-MAR-2003 (TREMBrel. 23, Created)				RX MEDLINE=2013006; PubMed=10731132;
DT 01-OCT-2003 (TREMBrel. 23, Last sequence update)				Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;				Amanatides P.G., Scheer S.E., Li P.W., Hoskins R.A., Galle R.P.,
OC Ephydioidea; Drosophilidae; Drosophila;				George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
OC Drosophila melanogaster (Fruit fly);				Sutton G.R., Worrell J.R., Yandell M.D., Zhang Q., Chen L.X.,
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;				Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champine M., Pfeiffer B.D.,
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;				Wan K.H., Doyle C., Baxter B.G., Heit G., Nelson C.R., Miklos G.L.G.,
RA				

RA Abril J.-F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,  
 RA Bellew R.M., Basu A., Baxendale J., Bayrktaroglu L., Beasley E.M.,  
 RA Besson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolhakov S.,  
 RA Botkova D., Botchkina M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo B., Deicher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Duigan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferriz C., Ferreira S., Fleischmann W.,  
 RA Fosler C., Gabriele A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegnami C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Misshina N.V., Moharry C., Morris J., Mosrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Neilson K.A., Nixon K., Nusskern D.R., Pacieb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reiner K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Sun E.,  
 RA Svirsks R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.-S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Venner J.C.;  
 RT "the genome sequence of *Drosophila melanogaster*";  
 RL Science 287:2165-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N\_A.  
 RC  
 RA Stapleton M., Brookstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise B., George R.,  
 RA Gonzalez M., Guaina H., Li P., Liao G., Miranda A., Mungall C.J.,  
 RA Nunez J., Paclob J., Parada V., Park S., Phouanenavong S., Wan K.,  
 RA Yu C., Lewis S.E., Rubin G.M., Celinker S.;  
 RL Submitted (OCT-01-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AR00355; AAF5621.1; --;  
 DR EMBL; AV058725; AAU19541; --;  
 DR FlyBase; FBgn0012643; CG6453.  
 GO; GO:0005059; F:calcium ion binding; IPI.  
 DR InterPro; IPR00048; EF-hand.  
 DR InterPro; IPR000886; ER\_target\_S.  
 DR InterPro; IPR00172; LDD\_receptor\_A.  
 DR InterPro; IPR00311; Man\_6\_P\_R\_bind.  
 DR InterPro; IPR00316; erhand; 1.  
 DR InterPro; IPR00317; ER\_HAND.  
 DR SMART; SM00192; IDLA; 1.  
 DR PROSITE; PS00018; IDLA; 1.  
 DR PROSITE; PS00014; ER\_HAND; 1.  
 DR PROSITE; PS00014; ER\_TARGET; 1.  
 DR PROSITE; PS50068; LDRA2; 1.  
 SQ SEQUENCE 54 AA; 61539 MW; 4F486B724D64732E CRC64;

Query Match 8.6%; Score 141; DB 11; Length 521;  
 Best Local Similarity 31.0%; Pred. No. 0.0011; Mismatches 53; Indels 12; Gaps 5;  
 Matches 36; Conservative 15; Mismatches 53; Indels 12; Gaps 5;

Qy 60 GPVHLER-LSGKGSFLVESTYKVECPFHNTQHEQPFMRNAYSG--LIGIHWHE-IA 114  
 Db 396 GPSGEFAVLYSQCVELTNEYVYRLCPPKLQYSQPK-----HGGSPTSLGWSWAGPD 449

Qy 115 NNTFTGMMRDGDAc-RSRRSQSKVELAGCSNRHLAHVSERSTCYTALETFPLVC 169  
 Db 450 HDKPSAMKVEQGTCWQPNRSTIVTRILCGKETVVTSTTEPSRCYELMTPAAC 505

RESULT 14  
 Q921X2  
 ID Q921X2 PRELIMINARY; PRT; 528 AA.  
 AC  
 DT 01-DEC-2001 (TREMBrel, 19, Created)  
 DT 01-DEC-2001 (TREMBrel, 19, Last sequence update)  
 DT 01-OCT-2003 (TREMBrel, 25, Last annotation update)  
 DB Similar to protein kinase C substrate 80K-H.  
 GN PRKCSC\_H  
 OS *Mus musculus* (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OK NBL1\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N\_A.  
 RA Strausberg R.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; BC005816; AAH05816; 1; --;  
 DR MED; MCI:107877; Pkcsch.  
 DR GO; GO:0010177; C:alpha-glucosidase complex; IPI.  
 DR GO; GO:000558; F:alpha-glucosidase activity; IPI.  
 DR GO; GO:000515; F:protein binding; IPI.  
 DR InterPro; IPR00048; EF-hand.  
 DR InterPro; IPR000886; ER\_target\_S.

DR	InterPro; IPR02117; LDL_receptor_A.
DR	InterPro; IPR00011; Man_6_P_R_bind.
DR	Pfam; PF00036; efbard; 2.
DR	SMART; SM00192; LDL; 1.
DR	PROSITE; PS00018; EF_HAND; 1.
DR	PROSITE; PS00044; ER_TARGET; 1.
SQ	SEQUENCE 528 AA; 59559 MW; 253PDIAT6A641713 CRC64;
Query Match	8.6%; Score 141; DB 11; Length 528;
Best Local Similarity	31.0%; Pred. No. 0.0011;
Matches	Mismatches 53; Indels 12; Gaps 5
QY	60 GPVHLR-LSKCFSIESTVYKVERPPHNVTQHQTFRNAYSG--ILGIWHEWAIA 114
Db	403 GPSGERPLVYDOSQVYVYVRLPPKLNQSKR-----HGSQPSLGTGKSWAGD 456
QY	115 NNTFTGMMRDGDAC-RSRRSQSKVFLAGKSNSRLAHVSEPTCVALTFTETPLVC 169
Db	457 HDKFSAMKYEQQGCGCQGPNSSTTVLLCGKETVWSTPSRCSYLMELMTPAC 512
RESULT 15	
Q96BU9	PRELIMINARY; PRT; 200 AA.
ID	Q96BU9
AC	Q96BU9;
DT	01-DEC-2001 (T-EMBLel: 19, Created)
DT	01-DEC-2001 (TRIMBLEl: 19, Last Sequence update)
DT	01-OCT-2003. (REMBLEl: 25, Last annotation update)
DE	Hypothetical protein (Fragment).
OS	Homo sapiens (Human).
OC	Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Buteleostomi;
OX	NCBI_TaxId:9605;
RN	11]
RP	SEQUENCE FROM N.A.
RC	TISSUE-LUNG?
RA	Strausberg R.;
RL	Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR	EMBL; BC015154; AAH15154_1; -.
DR	Intertroll; IPR00886; ER_target_S.
DR	Intertroll; IPR009011; Man_6_P_R_bind.
DR	PROSTTB; PS00044; ER_TARGET; 1.
KW	Hypothetical protein.
FT	NON_TER 1
FT	SEQUENCE 200 AA; 22628 MW; B9FA34051267DE9B CRC64;
Query Match	8.6%; Score 140.5; DB 4; Length 200;
Best Local Similarity	30.5%; Pred. No. 3.3e-05;
Matches	Mismatches 54; Indels 5; Gaps 3
QY	67 LSGKCSLIVESIYKVYKFCPPHNTVHQHTQRNAYSGILGIGIWHW-BIANTFTGMMD 1255
Db	83 LYSQKLTTEVYVYVRLCPKLVSOP--KLGSQSPTLSGTGWSWGPBDKESAMKYEQ 1399
QY	126 GDAC-PRSRSQSKVFLAGKSNSRLAHVSEPTCVALTFTETPLVC 169
Db	140 GTGCWQGPNSRSTVRLPCKETMTWTTESRCVLMELMTPAC 184

This Page Blank (uspto)